ARMY RESEARCH LABORATORY

Technologies for Army Knowledge Fusion

by Richard Scherl and Dana L. Ulery

ARL-TR-3279   **September 2004**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5067

Technologies for Army Knowledge Fusion

Richard Scherl
Department of Computer Science
Monmouth University

Dana L. Ulery
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
September 2004	Final	1 June 2003–31 December 2003

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Technologies for Army Knowledge Fusion	DAAD19-02-D-0001-0262
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Richard Scherl* and Dana L. Ulery	61102 T14
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Monmouth University Computer Science Department West Long Branch, NJ 7764-1898	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Research Laboratory ATTN: AMSRD-ARL-CI Aberdeen Proving Ground, MD 21005-5066	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
	ARL-TR-3279

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Department of Computer Science, Monmouth University, B 23 Howard Hall, West Long Branch, NJ 07764-1898

14. ABSTRACT

The fusing of information to provide a real-time situational awareness and common relevant operational picture of events both within and outside the battlespace requires the integration and use of heterogeneous civilian and military sources. Knowledge fusion, also called information fusion and multisensor data fusion, names the body of techniques needed to accomplish this critical need of the U.S. Army. The area builds upon and integrates results from a variety of fields, including multiagent systems, information integration, the semantic web, and the classical (mathematically based, engineering in orientation) fusion methodologies. The goal of work in knowledge fusion is to intelligently fuse massive amounts of heterogeneous information (not only from sensors, but also from databases, files, web pages, etc.) into a form which may be used by humans to yield actionable knowledge. This report surveys the literature relevant to the evolving field of knowledge fusion.

15. SUBJECT TERMS

knowledge fusion, multisensor data fusion, information integration, semantic web, intelligent agents

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Richard Scherl
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL	56	19b. TELEPHONE NUMBER *(Include area code)* 410-278-8609

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

ii

Contents

1. Introduction

1.1 Motivation

The U.S. Army has identified a critical need to fuse information horizontally across multiple peer units and vertically up and down the chain of command to create real-time and near-real-time knowledge at all levels of U.S. Army operations. The concept of fusing information from a wide variety of battlefield assets has been evolving for the last 20 years to improve command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) capabilities for tactical wartime and contingency operations, situational assessment and awareness, and real-time collaboration. Most recently, the science and technology (S&T) initiatives identified to enhance the U.S. Army's ability to accomplish its emerging Homeland Security mission has underscored the need to fuse multisource information to provide a Common Relevant Operational Picture built on shared data.

We use the term "knowledge fusion" to emphasize the real-time integration and use of information from heterogeneous civilian and military sources to assist the human process of knowledge creation. This concept builds on the emerging field of multisensor data fusion technology. Multisensor data fusion focuses on the integration of data from multiple sensors to achieve more accuracy than could be achieved by using a single, independent sensor (Hall and Llinas, 2001). The Joint Directors of Laboratories (JDL) Data Fusion Working Group was established in 1986 to improve communications among researchers and developers. The two-layer, hierarchical JDL process model that resulted from its efforts identifies the processes, functions, and techniques applicable to data fusion. The 1998 model revision broadly defines data fusion as "the process of combining data or information to estimate or predict entity states" (Steinberg et al., 1998). Excellent overviews of multisensor data fusion are provided by Hall and Llinas (2001), Waltz and Llinas (1990), and Antony (1995).

Knowledge fusion also builds on the emerging fields of multiagent systems (Subrahmanian, et al., 2000; Genesereth and Fikes, 1992), information integration (Wiederhold, 1992, 2002), (Jhingran et al., 2002); and the semantic web (Fensel, 2001; Fensel et al., 2003a). Knowledge fusion focuses on the integration, system techniques, and principles that increase the value of information produced by integrating information from multiple real-time and non-real-time heterogeneous sources that include, but are not limited to, sensors.

In the battlespace, information is continuously arriving from sensors of various types on different platforms. The output of these sensors along with verbal reports, electronic mail and reports, background knowledge, current breaking web-based news reports, and other types of information need to be fused into a coherent picture of the immediate situation so that the best courses of action may be determined. In practice, the terms "knowledge fusion" and "multisensor fusion"

are both commonly used to refer to advanced technologies for integrating data from multiple sources for rapid and effective decision making. However, areas of focus and techniques for solving the problem tend to stem from different traditional and emerging disciplines.

Knowledge fusion information sources may store data in various formats within databases, files, or on web pages. Alternatively, the data may be the real-time streaming output of various devices. Furthermore, the structuring or categorization of the information may be radically different in the various sources. Additionally, the time varying nature of different types of information may be different. Some data are relatively static while other sources are continuously changing. The goal of work in knowledge fusion is to develop techniques to intelligently fuse and process massive amounts of heterogeneous information from a wide range of distributed sources so that it may be used for decision making and problem solving.

Knowledge is emerging as the ultimate in force multipliers. As information technology advances in capabilities and security, the symbiosis between machine and human increases in an unprecedented way. A core capability in the concept of knowledge fusion is support of human knowledge processes, accomplished through simplification and filtering techniques, and multimedia presentation techniques that can be tailored to the preferences and skills of the human receiver. Technologies to support these capabilities in military domains differ from those in civilian domains because systems on the battlefield and in recovery and consequence management (R&CM) situations must accommodate a real-time, on-the-move, dynamic, and extremely high-stress environment with common operations performed by humans with a wide range of skills and tasks.

1.2   Discussion of Terminology

The words "data," "information," and "knowledge" are used in varying and overlapping ways in not only everyday discourse, but also in the scientific literature. There is the rapidly evolving area of data fusion which is primarily concerned with mathematical techniques to combine the output of sensors. The most mature area of multisensor data fusion is aimed at combining sensor data to determine the position, velocity, attributes, and identity of individual objects to identify and maneuver targets. Within the computer science field of databases, the well-established area of data and database integration addresses combining data sources. This field studies techniques to combine information from databases with different schemas, but also has been extended to incorporate information from less-structured sources such as files and web pages. The area of knowledge fusion includes both of these areas, but also defines and addresses processes that increase the value of information achieved by combining and relating information from a wide variety of sources. We do not expect to combine materialized data sources, but rather to bring together only selected results derived from these sources. We focus on the human problem of data overload caused by the current explosion of battlespace data from local and remote sensors, human voice and email input, web-based data, and multimedia news data. The knowledge fusion

problem is finding and putting together pertinent information relevant enough to make informed decisions and take swift action when confronted with a mountain of data.

There is some philosophically oriented literature on the relationship between the notions of data, information, and knowledge (Dretske, 1999; Devlin, 1991). Although, it is not our intention to prescribe the way these terms should be used, a mention of this discussion does lend some clarity to the varying approaches to be discussed in this report, particularly since the philosophical discussion does at least, in a very loose sense, correspond to the way these terms are generally used. In ordinary usage, there is certainly a continuum from data to information and then to knowledge. The notion of data is the most primitive (e.g., raw data). The notion of information implies more interpretation and processing. Finally, knowledge carries the implication of general principles or facts of a propositional form, which are known to be true.

Devlin (1999), drawing upon the work of Dretske (1999), situation semantics (Devlin, 1991), and also business researchers such as Davenport and Prusak (1998) argues for the utility of distinguishing between the terms in the following manner. Information is data along with meaning. And knowledge is information which has been internalized and associated with the ability to use the information. Knowledge is then an individual notion as it exists within a person. Similar views are expressed by Brown and Duguid (2000).

The approach of situation semantics aims to establish a Mathematics of Information (Devlin 1991, 1999; Barwise, 1989) by further refining the view of information as data along with meaning. The core idea is that information is made up of some form of representation and a procedure for encoding and decoding the information. The encoding/decoding procedure is based upon constraints between the representation and the information conveyed about the world. An important outcome of this approach is the idea that information can only be understood or decoded if contextual situations are identified (Devlin, 1999). The issue of representing and maintaining the context (e.g., temporal, geographical, terminological, etc.) in which data/information are gathered, and utilizing that contextual information to combine various sources into a unified picture, is one that in various forms underlies much of the work discussed in this report.

1.3   Outline of Report

This report identifies and analyzes technologies relevant to U.S. Army knowledge fusion from a variety of fields such as multisensor data fusion, intelligent agents, semantic web, information integration, databases, artificial intelligence, systems engineering, and others. The goal is to identify and evaluate techniques that have high-potential applicability to Army knowledge fusion problems.

In section 2, we describe the U.S. Army's Future Force, the vision of the army of the future. The aim of knowledge fusion is to integrate critical information to enable knowledgeable action required by the Army's Future Force. Section 3 describes some examples of Army knowledge

fusion issues. Technologies relevant to knowledge fusion are identified and discussed in section 4. Finally, technologies of importance to the Army knowledge fusion problem are analyzed and summarized in section 5.

2. Background

The U.S. Army's Future Force (Department of the Army, 2004; Williams, 2003; U.S. Army Training and Doctrine Command, 2002), the army of the 21st century, is being designed to be joint, interagency, and multinational (JIM). Land power plays a critical role in dominating the highly complex land environment that comprises the heart of most joint operations. There is an enduring and unavoidable challenge to control terrain, people, and resources on land, where political authority resides. The Future Force is to be an instrument that is precise, maneuverable, and flexible. It must be able to fight on a battlefield that is multidimensional, dispersed, continuous and noncontiguous in nature.

The units to be deployed must be tailored to the operation at hand. The units will incorporate the appropriate equipment and personnel for the operation drawn from multiple U.S. agencies as well as those of other countries. The operation of the units are network-centric in that commanders in different places can obtain relevant information.

A crucial role is to be played by the distributed information environment called the Global Information Grid (GIG), which is to provide links from the "factory to foxhole" and "space to mud." The architecture of GIG needs to be designed to create a knowledge-based force in which soldiers receive and send the right information at the right place and time. This information environment is to facilitate the transformation of data into knowledge.

In this section, the demands of the battlespace, sustaining the base, and also homeland security are discussed in the context of the Army's Future Force and GIG.

2.1 Battlespace

The Army's Future Force, supported by GIG, will carry out Network Centric Warfare (NCW) which is defined as

> an information superiority-enabled concept of operations that generates
> increased combat power by networking sensors, decision makers, and shooters
> to achieve shared awareness, increased speed of command, a higher tempo of
> operations, greater lethality, increased survivability, and a degree of
> self-synchronization (Alberts et al., 2002).

Through the linking of knowledgeable entities in the battlespace, NCW is to convert information superiority into combat superiority. The backbone of this information system is an integrated

communications system. This includes both wireless and wired networking. The network connects all battlespace entities—sensors, actors, and decision makers.

In the network-centric model, sensors and actors are decoupled, unlike the traditional platform-centric model. Sensors from one platform may provide information that enters the network and is ultimately used by a variety of other actors and decision makers. Additionally, new types of sensors are available and need to be effectively used.

> Sensors designed to sense new things and maneuver in close to make distinctions among things we cannot now distinguish—The net result of all of these changes will be the proliferation of lower cost, independent sensors and actors that will contribute to and depend more upon distributed rather than embedded intelligence (Alberts et al., 2002).

Information collected by the sensors needs to be put in a form so that it can be used by other battlespace entities. The other entities need to fuse the information, interpret it in context and understand the implications (Alberts et al., 2002). Note that the knowledge/information fusion issue arises immediately here because sensor information from a sensor designed for use by a particular battlespace entity is being made available for wider use. Additionally, we also see the importance of movement or positioning for the purpose of information gathering.

The ultimate goal of this fusing is the construction of common relevant operational picture (CROP) (U.S. Joint Forces Command, 2004; Falvo, 2004). The CROP will provide a presentation of timely, fused, accurate, and relevant information that can be tailored to meet the requirements of the Joint force commander and the Joint force, and is common to every organization and individual involved in a joint operation. A CROP facilitates collaborative planning and assists all echelons to achieve situational awareness.

> Providing battlespace awareness to war fighters across the Joint force with requisite accuracy and timeliness requires that data and information from multiple sources be collected, processed (analyzed when necessary), transported, fused, placed in appropriate contexts, and presented in ways that facilitate rapid and accurate inferences. It also requires that actors and decision entities be provided by training with internal models and or decision aids or models. With this insight, we can observe that it requires both battlespace awareness and these cognitive models to generate battlespace knowledge which is in and of itself, an emergent network-centric property (Alberts et al., 2002).

The integrated information found in the CROP includes (Alberts et al., 2002) the following:

1. The location of the various friendly forces, enemy sources, and other entities.

2. The available courses of action for both friendly forces and enemy forces.

3. The predicted future actions and positions for all forces.

4. The status and characteristics of various units.

5. The environment, including the terrain and weather conditions.

Battlespace awareness results from the fusion of key types of information that describe the battlespace. These include positions of forces and geography. But battlespace knowledge consists of tacit information that requires interpretations. Examples are intent and capabilities, and tactics of the various forces. The information used to build the CROP will come from a variety of sources. It will not only include the battlespace sensors but also, for example, the information available on news networks such as CNN.

Yet another aspect of NCW is to capitalize upon the fact that the actor entities are knowledgeable. Command and control (C2) may be replaced by command and coordination. Thus the desired results may be achieved without the issuance of top-down detailed instructions.

> That is to say that organizational behavior could be consciously designed to be an emergent property that derives from the commander's intent, as internalized by actor entities, the degree of battlespace knowledge available and the ability of decision entities to minimize the constraints imposed on actor entities by virtue of the resources allocated to actor entities. It is hard to overestimate the impact that this new dimension of command and control will have on the way we will approach operations in the future (Alberts et al., 2002).

How to facilitate and control the execution of such a loose and distributed plan or task is an open question.

Additionally, it is expected that use will be made of battlespace agents, which perform selected tasks as delegated by the decision and actor entities. We not only have links between sensors, actors, and decision entities, but battlespace agents also play a role. These agents are automated decision or information processes (Alberts et al., 2002).

Some of the tasks that may be given to battlespace agents are as follows:

1. Request additional information that is required by the situation.

2. Task the sensors.

3. Notify the decision entities that something requires immediate attention.

4. Translate a commander's intent into instructions.

5. Resolve inconsistencies within the CROP.

In summary, the vision of the future U.S. Army reveals the crucial importance of knowledge and information fusion in the information environment needed to support the Joint operational activities. Massive amounts of continuously arriving information from a variety of sensors and information sources, in a variety of formats and from various locations and perspectives, must be

combined into a unified picture that is able to discern and present only that relevant set of information that the warrior needs to make informed decisions and take appropriate action. The fact that the CROP is available to a variety of battlespace agents produces the potential for a new form of C2 through coordination rather than exclusively and centrally directed control. Automated agents and robots are to be well integrated into this networked system in a way that results in a symbiotic relationship between warrior and machine.

2.2   Sustaining Base

Some issues found in the commercial sector also arise in the Army's sustaining base domain, but with significantly different constraints and requirements due to the fact that the sustaining base supports a rapidly changing battlesphere with high human risk factors.

Inventory is stored at various locations, and as it is used, orders need to be made to replenish the supply. Information on the nature of the items stored in various locations is kept in different formats on various types of databases. While as-needed work is being carried out, information systems need to ensure that all the required steps (e.g., maintenance operations) are performed.

One major difference between the military case and the commercial sector is the interaction with the battlespace. For example, equipment must be stored at appropriate locations in anticipation of events in the battlespace, and events in the battlespace will initiate supply efforts on the sustaining base. The events of the battlespace are more difficult to predict than those that occur in the private sector. It may, for example, be necessary to move large numbers of soldiers from one location to another with very short notice and have all the necessary arrangements (e.g., food, housing, equipment, transportation) made.

Furthermore, the fact that the Army's Future Force is to draw equipment and personnel for operations from multiple U.S. agencies as well as those from other countries means that the information integration task will be immense. It is to be expected that the terminology and categories used to describe available equipment and personnel will be radically different between the Army and the various U.S. and foreign agencies with which efforts need to be coordinated.

Finally, delay and error can result in loss of life. Information concerning the status of sustaining base operations can assist the enemy, with potential negative impact of the operation that again includes resulting loss of life.

In summary, some sustaining base knowledge fusion tasks have similarities with those found in more traditional information integration tasks that arise in the commercial sector. But the military domain augments the challenges of these fusion tasks considerably and adds other critical dimensions because of the rapidly changing nature of the battlespace, the high need for security, and the significant impact of delay and error.

2.3 Homeland Security

As is now well known, prior to 9/11/01, there was information available which could have possibly alerted us to the impending attacks. The question arises as to whether an advanced information system could have put all the pieces together, predicted the attack, and allowed countermeasures to be taken.

There is an immense amount of data available; some in public sources and other bodies of data explicitly gathered by different governmental agencies. The number of possible types of threats is immense. Additionally, these can occur at many different locations and at any time. Issues of knowledge and information fusion arise in putting the different pieces together to get a unified picture of where the likely threats are.

Again, one sees the traditional information integration issue here. But the amount of data is vast and continuously changing. There is really a knowledge fusion problem beyond that of information integration. The question is how can the crucial pieces of information be identified (from among the vast amounts of unimportant information) and put together so that action can be taken.

2.4 Some Issues

Certainly, traditional information integration issues concerning how to combine information from heterogeneous structured (e.g., databases) and unstructured (e.g., files and web pages) information sources is of importance to addressing the knowledge fusion issues that are to arise in the U.S. Army of the future. Given the nature of the Army's Future Force, i.e., the utilization of carefully tailored units with membership drawn not only from the U.S. Army, but also from foreign military forces as well as various U.S. Governmental agencies, there will be a particularly massive information integration problem.

Given the nature of NCW, the amount of data arriving from all available sensors will be massive and continuously arriving. The techniques of data fusion are certainly needed. Given the number of sensors available, the issue of how can one automatically determine when a change in the data (continuously streaming data) is significant. Additionally, in NCW, the sensors used by a particular actor are not at the same position as the actor. The information from the various sensors need to be combined to give the knowledge/information needed by the actor to carry out its task(s) and achieve its goals. Note that sensors may be moved to particular locations (independently of the actors) in order to obtain information that would satisfy particular knowledge goals. These knowledge goals are necessary to satisfy the overall battlespace goals. It is necessary to represent knowledge goals and develop techniques for obtaining plans to achieve the knowledge goal.

There is an added issue of the representation of plans. When a crucial piece of information arrives, indicating that a particular event has occurred, how is it possible to identify the plans or ongoing activities that are affected by this event? What changes then need to be made?

Additionally, it is expected that knowledgeable actors can self-synchronize rather than being directed from the top. How is it possible to represent the ongoing plan or activity in such a way that this self-synchronization is possible?

There may be a wide variety of automated tools or agents available on the network. Somehow, these need to be categorized and set up in such a way that the particular services that they provide can be determined either automatically or by a human user. Techniques for doing so need to be identified and/or developed.

A number of areas of computer science that are certainly relevant to the Army's Future Force and NCW have not been included. These are as follows:

1. Issues related to human-computer interfaces and human factors relevant for the design of CROP.

2. Computer/network security issues.

These issues are certainly important, but the literature related to these issues is vast and outside the scope of this report.

3.   Motivating Examples

It is useful to have some detailed examples of knowledge fusion issues that might arise in the operations of the U.S. Army to facilitate work on this topic. This section describes several places in the literature where the beginnings of such scenarios are found.

3.1   Movement Analysis Battlespace Challenge Problem

The Defense Advanced Research Projects Agency (DARPA) High Performance Knowledge Bases (HPKB) Project used a number of problem types and specific scenarios (Cohen et al., 1998). One of these was the Movement Analysis Battlespace Challenge Problem.

The goal of this problem was to understand sensor data pertaining to the movement of vehicles. The sensor data were an idealized version of that produced by airborne Joint Surveillance Target Attack Radar System (JSTARS). This was supplemented with some intelligence reports concerning when enemy radar has been turned on, information on enemy communications and human intelligence. Information on the geography, roads, and the composition of enemy forces in the region was also given.

The task of the movement analysis involved the following:

1. Distinguish military from nonmilitary traffic.

2. Identify the sites between which the military traffic is moving and determine their types (e.g., battle position, support area, artillery site) and significance.

3. Identify which units are participating in each movement (convoy).

4. Determine the purpose of the convoy (e.g., reconnaissance, engage in battle, etc.).

5. Determine the exact types of the vehicles.

See Cohen et al. (1998) for more details.

Solutions to this type of scenario would directly serve the needs of the Army's Future Force. The construction of the CROP in network centric warfare calls for this kind of interpretation of enemy actions.

3.2 Workaround Battlespace Challenge Problem

The purpose of this problem, also from DARPA HPKB Project, was to devise a system to help determine targets for destruction by determining how rapidly the enemy can bypass (workaround) the damage to the target (Cohen et al., 1998). Although, the researchers in the HPKB project had primarily air strikes in mind, the same sort of example can be used to reason about the choice of targets for destruction or capture by ground forces.

Input to the task includes the following:

1. Description of the target, the damage to it, and key features of the geography of the area.

2. The purpose of destroying the target, e.g., to prevent a particular enemy unit from moving to a particular position. Additionally, the input includes the time period during which the enemy capability needs to be interdicted.

3. Description of the enemy resources.

The output of the workaround generator produces a reconstitution schedule, giving the capacity of the enemy to either repair the target or to find another way (workaround) of accomplishing the activity mentioned in point 2. The workaround generator needed substantial information about the engineering capabilities of the enemy. Again, solutions to this type of scenario are of direct relevance to the problem of supporting the Army's Future Force. Kingston (2001) contrasts several approaches to this problem.

3.3 Joint Warrior Interoperability Demonstration 2003

Another example is the Joint Warrior Interoperability Demonstration 2003 scenario (Office of the Army, 2004). The scenario is set in an area of operations deep in the Pacific Ocean, centered on a fictitious island called Tindoro. Tindoro is divided due to a turbulent political history. The North, the opposition force, has created a province that is politically and economically attached to the mainland. The South is an independent nation with friendly relations with a nation called

Rabenneste. It controls the Fingal Enclave located within the North territory. As the action begins, the region has been destabilized to the point of needing United Nations (UN) intervention. A conflict occurs for a period of 220 days. The UN inserts substantial force into the South, and captures Flinders Island. The North attacks Fingal Enclave. The UN Multinational Task Force (MTF) recaptures the Enclave. Peace Keeping Operation forces enter the South and provide humanitarian aid. MTF consolidates forces in the South and enforces peace keeping agreements. This scenario demands the ability for 43 coalition partners in wide-spread, multiple geographical locations to share relevant operational information reliably and quickly. These nations are not necessarily linked by treaty. MTF staff officers must have secure and reliable access to near real-time situational awareness.

3.4   Homeland Security Problem

A good example for homeland security is exactly what happened prior to 9/11/01. The recently released Congressional Report contains much material. The question is what type of information system would have put things together and issued warnings. The data would include the Phoenix Report from July of 2001, a memo from an FBI agent in the Phoenix field office indicating that attention should be paid to individuals from the middle east who have enrolled in U.S. flight schools. Additional information would be money transfers such as that from Ramzi Bin al-Shibh (who is assumed to be a member of a German al Qaeda cell) to Zacarias Moussaoui who was enrolled in a flight school in Minnesota. This type of data and the various sources of information, where it was available, could be developed into a testbed problem for knowledge fusion efforts.

4.   Technologies

Technologies relevant to constructing the information system needed to support Army knowledge fusion are quite broad, covering many areas from multiple disciplines, such as Sensors, Databases, Artificial Intelligence, Systems Engineering, etc. We have divided up the literature into manageable units. To some extent, the determination of units is arbitrary and others may prefer a different division. But nevertheless, we feel that the categorization chosen here helps in understanding the wide variety of techniques which are available to address the knowledge fusion problem. Sections 4.1 and 4.2 contain the relatively well-established areas of Data Fusion and Information Integration. The two areas of Ontologies and also the Semantic Web are grouped together in section 4.3. Section 4.4 contains the somewhat disparate set of topics grouped together under the label Advanced Representation and Reasoning. These include planning, plan recognition, logic programming, Bayesian networks, and real-time problem solving. A number of topics related to databases did not properly fit into any of the earlier sections. These include research on querying streams, querying unstructured information sources, web mining, and data mining. They are in section 4.5, labeled Contemporary Database

Topics. Finally, in section 4.6, topics related to the design of the overall architecture of the information system to support the Army's Future Force are covered.

4.1 Data Fusion

Data fusion is a name for work carried out primarily with military applications in mind. The typical problems are object identification, combining information from multiple sensors, object tracking, threat assessment, and so on. The methods utilized by work in this area are generally numerical and mathematical techniques based on linear algebra, probability theory, stochastic processes, and statistics. The overall problem has been defined as follows:

> Locate and identify an unknown number of unknown objects of many different types on the basis of different kinds of evidence. This evidence is collected on an ongoing basis by many possibly re-allocatable sensors having varying capabilities. Analyze the results in such a way as to supply local and over-all assessments of the significance of a scenario and to determine proper responses based on those assessments (Goodman et al., 1997).

As noted by Goodman et al. (1997), this definition could apply to the concerns of other areas such as pattern recognition and related aspects of artificial intelligence, e.g., those described in Duda et al. (2001).

In the typical military setting, there are various platforms with sensors of different types. From such information, the data fusion system needs to produce reasonable hypotheses concerning the actual truth on the ground. How many vehicles are there? What are their positions? What are their velocities and directions of movements? What are their identities and military status (enemy, allied, or neutral)? The amount of data is vast and continuous. Various mathematical techniques have been developed to deal with this problem.

These techniques include the Kalman filter, Multi-Hypothesis Estimation (MHE) filter, Joint Probabilistic Data Association (JPDA) filter, the Mori, Chong, Tse, and Wishner (MCTW) filter, Point-Process Filter, Hidden Markov Model (HMM) Filters and many others as well. Additionally, the techniques include the use of Belief Functions (Dempster-Shafer Theory of Evidence), random sets, event algebras, neural networks, and also Bayesian techniques (Robert, 2001; Gelman et al., 2000). Simulation and optimization techniques are also used, i.e., to determine sensor placement (Brown and Schaumburg, 2001). The literature is vast and a detailed comparison of the various techniques cannot be given here (see Hall, 1992; Waltz and Llinas, 1990; Sadjadi, 1996; Bar-Shalom and Fortmann, 1998; Klein, 1993).

Some recent efforts towards improving these techniques by adding world/domain knowledge are due to Donald Brown and his collaborators (Power and Brown, 2002; Sobiesk et al., 1998; Barker et al., 1998; Bovey and Brown, 2000). The types of knowledge that they incorporate include knowledge of the environment (e.g., terrain) and also characteristics of the entities operating in the environment (e.g., membership in a group such as a convoy). Improved

performance results from incorporating this additional information. This work is significant in that it forms a bridge between the traditional data fusion activities and the higher-level techniques to be discussed next.

Donald Brown and his collaborators have also used the techniques of data fusion to predict criminal incidents on the basis of past occurrences and to determine which crimes are likely to have been committed by the same individuals (Brown, 1998; Liu and Brown, 1998; Brown and Hagen, 1999; Brown and Oxford, 2001; Xue and Brown, 2003). Essential to this work is a mathematical measure of the similarity of incidents. Clearly the same techniques can be applied to the prediction of incidents such as terrorist attacks or military strikes.

A general model of data fusion, the Joint Directors of Labs (JDL) model, is laid out by White (1988) and then revised by Steinberg et al. (1998). The later paper defines data fusion very generally as the process of combining data to refine state estimates and predictions. A number of levels of data fusion are proposed by Steinberg et al. (1998) to categorize the different types of problems addressed in data fusion work. The levels are as follows:

> Level 0: Sub-Object Data Assessment, estimation and prediction of signal/object observable states on the basis of pixel/signal level data association and characterization.

> Level 1: Object Assessment, estimation and prediction of entity states on the basis of observation-to-track association, continuous state estimation, and discrete state estimation.

> Level 2: Situation Assessment, estimation and prediction of relations among entities.

> Level 3: Impact Assessment, estimation and prediction of effects on situations of planned or estimated/predicted actions by participants.

> Level 4: Process Refinement, adaptive data acquisition and processing to support mission objectives.

Level 0 involves things like detecting signals and finding features in an image. Level 1 involves the identification of physical objects, possibly by drawing upon other fusion processes, i.e., tracks. Groupings of these objects are postulated under level 2. Level 3 aggregates or relates the groupings of entities from level 2 into plans and determines how they might effect the plans of other parties. Finally, level 4 determines what actions should be taken perhaps as part of a new plan. Steinberg et al. (1998) state clearly that the information flow may go in multiple directions, not just from level 1 to level 2 to level 3 and then to level 4. Additionally, they point out that particular applications may call for a different partition of functions.

Clearly, the JDL architecture is motivated by the integration of sensor data. It may not be immediately suitable for the integration of other sorts of information such as the contents of databases. Additionally, the higher up one goes in levels (from 2 to 3, and certainly to 4); the less applicable are the traditional mathematical methods of Data Fusion. Yet the overall goals

are not very different from that of Knowledge Fusion. The model seems to call for or need the techniques that will be discussed in the remainder of this report.

4.2   Information Integration

The goal of work in information/data integration is to provide uniform access to multiple sources of information. The idea is that the user should only have to focus on specifying what is wanted and to let the system worry about how to obtain the answers (Levy, 1998, 1999; Levy and Weld, 2000). Although initial work in this area considered only databases, the work was quickly generalized to include web sources as well.

Information/data integration work has been primarily carried out by database researchers. An introductory presentation of issues of information integration may be found in Chapter 20 of Garcia-Molina et al. (2002). Surveys of earlier work in this area include Batini et al. (1986) and the more recent work of Hull (1997). Wiederhold (1992) introduced the important notion of mediators, i.e., modules that mediate between the users' workstations and data sources, using knowledge to transform data into information.

The literature of the past decade describes a number of projects in this area:

1.   TSIMMIS (Garcia-Molina et al., 1997) is an information integration architecture based on mediators. The architecture includes the following:

   a.   The OEM data model, an object oriented data model that is self describing (Nestorov et al., 1997; Papakonstantinou et al., 1995). It is especially suitable for semistructured hierarchical data such as that found on the World Wide Web.

   b.   Mediators and wrappers and tools for generating them. A system for declaratively specifying mediators has been developed (Papakonstantinou et al., 1996). Papakonstantinou et al. (1995) describe a toolkit for the development of wrappers (software modules that extract data from a source).

   c.   A query language, LOREL, especially designed to query semistructured data or data from several heterogeneous data sources. The queries can operate over data having different types and also when some of the data is absent (Abiteboul et al., 1997).

The approach used in TSIMMIS has been described as "Global As View" (GAV)[*] because the mediated or global schema is described as a set of database views (i.e., queries) over the different information sources. The amount of computation needed to answer a query is thus minimized, but it is more difficult to add, delete, or modify information sources.

---

[*] Another term is query-centric.

2. SIMS: The SIMS project basically takes a planning approach to the problem (Arens et al., 1993, 1996). Given a query, the planner considers alternative ways to find the data required to satisfy the query.

3. InfoMaster: The approach used here is "Local As View" (LAV).* For every data source, it is necessary to describe the information in the source by writing a rule over the relations in the mediated (global) schema (Duschka and Genesereth, 1997a, 1997b). Query reformulation is therefore more difficult and requires a certain amount of inferencing. But adding additional data sources is easier.

4. InfoQuilt: (Thacker et al., 2003) is a platform for answering complex information requests over data available on the web. Ontologies (see section 4.3) are used to describe knowledge about domains and the relationship between them. InfoQuilt uses the LAV approach for representing the information contained by various sources. It has a number of tools to facilitate the construction of the knowledge base by the user.

5. HERMES: The HERMES system, developed by V. S. Subrahmanian and collaborators provides a particularly rich language for writing mediators (Subrahmanian et al., 1995; Subrahmanian, 1994; Adali and Subrahmanian, 1994, 1996; Lu et al., 1995, 1996; Adali and Emery, 1995). The language is based upon Annotated Logic Programming due to Kifer and Subrahmanian (1992).

6. Information Manifold: (Levy et al., 1996a, 1996b; Kirk et al., 1995). This work is designed to provide an interface to structured information sources available on the internet. The sources seem to be generally databases. The contents of sources are described in a language called CARIN (Levy and Rousset, 1996, 1998) that combines Datalog and a description logic (see section 4.3). The system computes a query plan to answer a particular query given the description of the available information sources. Probabilistic information on the coverage of data sources has been added (Florescu et al., 1997).

7. Nimble is a commercial system based upon XML (Draper et al., 2001) (see also http://www.nimble.com).

8. Tukwila: The Tukwila information integration system is designed to handle XML data and in particular streaming data. As such, Tukwila can provide answers while the data are continuously coming (Ives et al., 1999, 2000a, 2000b, 2001, 2003).

9. Softbot Interface to Internet: Etzioni and Weld (1994) describe a softbot interface to the internet which makes use of AI planning methods to satisfy the user's queries. Etzioni (1996) covers general issues concerning mining the Internet. Theoretical issues pertaining to constructing plans for information gathering when there are costs (monetary and

---

* Another term is view-centric.

15

temporal) on accessing different information sources are investigated by Etzioni et al. (1996).

10. Ariadne (Knoblock et al., 2002a) is a system for extracting and integrating data from web sources. It allows users to rapidly create information agents for the web.

The projects previously described have ranged from experimental to commercial. They have illustrated a variety of different techniques. But separate from particular projects such as those described previously, there has been quite a bit of work on general issues (both theoretical and practical or algorithmic) that arise in the area of information integration. These topics are surveyed in the following paragraphs.

Hull (1997) and Ullman (1997) survey approaches towards the integration of information from multiple databases. General work on the use of description logics for information integration has been carried out by Lenzerini and collaborators (Calvanese et al., 1997a, 1998a, 1998c, 2002). They developed a description logic (along with associated inference methods) that is suitable for representing the various information sources as well as the global or mediated domain. The logic can express the relationship between concepts from the different domains. Other applications are data warehousing (Calvanese et al., 1998b), handling SGML documents (Calvanese et al., 1997b), and extensions of object-oriented modeling (Calvanese et al., 1995). Additional general work on the use of description basics in information integration has also been carried out by Beeri et al. (1997).

General issues concerning query-answering plans for data integration systems are considered by Duschka, et al., (2000). In particular, they look at the notion of a maximally contained plan, a plan that provides all the answers that are possible to obtain from the available sources, but which may not necessarily be equivalent to the original query. The authors consider the cases when recursive plans are needed.

Friedman et al. (1999) discuss the use of planning techniques to gather information. The user's queries are translated into plans over information sources. The efficiency of information gathering plans and techniques for transforming plans into more efficient plans are considered in a number of papers (Friedman and Weld, 1997; Doan and HaLevy, 2002). Issues pertaining to query containment are considered by Milstein et al. (2002). Additionally, work at ISI has addressed the issue of planning queries in the presence of mediators that provide access to information distributed over heterogeneous sources (Ambite and Knoblock, 2000).

The question of optimizing query plans with regard to cost and coverage has been considered by a number of researchers. Nie and Kambhampati (2001) consider cost in terms of planning cost. This is traded off the coverage of the plan. Kambhampati and his collaborators have also looked at utilizing machine-learning techniques for gathering coverage statistics (Nie and Kambhampati, 2003; Nie et al., 2002).

In a number of publications, Bertossi, Chomicki, and collaborators look at the computation of consistent answers from inconsistent databases (Bertossi and Chomicki, 2003; Bravo and Bertossi, 2003; Arenas et al., 2000, 2003; Bertossi et al., 2002). They are particularly concerned with databases that violate integrity constraints. The approach is to compute the minimal repairs that could restore consistency to the database(s) and reason over this new set of databases.

Machine learning techniques have been proposed for learning many of the pieces of knowledge needed by information integration systems. Examples are learning the mapping between source and mediated schema (Doan et al., 2000), learning mappings between ontologies (Doan et al., 2002), and more generally using statistical methods for finding mappings between different schemas (HaLevy and Madhavan, 2003). Kushmerick (2000) analyzes a number of techniques for automatically inducing wrappers for web sites. Craven et al. (2000) developed a trainable system that extracts information from the web and produces a knowledge base.

Cohen (2000) describes a language called WHIRL that combines both textual (concepts based on information retrieval [I.R.]) and logic-based representation systems (e.g., datalog) into a single language for information representation. The language is useful for a variety of information integration tasks including the generation of wrappers.

Work in the area of information integration often makes use of the notions of mediators and wrappers. These are in some respect similar to the notion of agents. There are different pieces of software serving particular purposes, but it is a relatively passive model. The mediators and wrappers do not monitor their environment and take appropriate actions given their goals. For time varying data, it seems likely that the more active model of agents would be appropriate rather than the more passive model of mediators and wrappers. For keeping the COP up to date, sources will have to notify the COP of significant changes. How they would determine which changes are significant is an open question at this point. There may have to be extensive communication going on in both directions between the CROP and the wrappers.

4.3   Ontologies and the Semantic Web

This section covers several related areas. There has been an interest in ontologies, a machine-understandable specification of a conceptualization of a domain (Gruber, 1993), for some time in the area of artificial intelligence (AI)/knowledge-based systems. A number of languages have been developed to represent ontologies. The effort known as the semantic web has moved this work onto the web (using XML) so that the terminologies defined by these ontologies can now categorize data available on the web.

4.3.1 Ontologies

Overviews of recent work on ontologies have been written by Fensel (2001) and McGuinness (2003). For a more philosophical perspective, see Guarino (1998). Of particular recent importance is the convergence between web technology and ontologies. This convergence has resulted in efforts to create ontology languages based upon XML and RDF. The construction

and manipulation of ontologies are crucial to the development of the information system to support NCW and the Future Force. Ontologies provide a formal specification terminology used to categorize objects and events. Without such ontologies, it is simply not possible to automatically manipulate the terminologies so defined for fusion tasks.

4.3.2 Knowledge Representation Systems/Languages

There are a number of knowledge representation languages that have been developed to represent ontologies. Some of the better known are as follows:

1. Ontolingua is a frame-based language designed to facilitate the use of ontologies within multiple applications.

2. Loom was developed at ISI and is based on a description logic or classifier system.

3. See Reed and Lenat (2002) for a recent discussion of applications of CYC. See Lenat and Guha (1989) for the original description of the project. The CYC project includes both the development of an ontology language, a set of tools, and also a large ontology designed to be used in a wide variety of applications.

4. GKB (Generic Knowledge Base) Editor (developed at SRI) is a graphical tool for interactive knowledge base browsing and editing (Karp et al., 1998).

Karp (1993) has produced a (now somewhat dated) comparison of many of these systems.

Much work in the area of ontologies is based upon frame languages. A more flexible alternative that automatically computes the classification of instances is description logics. The literature on description logics is quite large. See Calvanese et al. (1998) for an overview of their use in modeling. A book length overview of the entire area of description logic research is now available (Bader et al., 2003).

Some work, especially at ISI, has been done on developing tools to assist in the gathering of knowledge (Blythe et al., 2001; Blythe, 2001; Gil and Ratnakar, 2002). Some of the projects are as follows:

1. Trellis (Trellis, Information Sciences Institute, USC [2004]; Gil, 2003).

2. Expect (Expect, Information Sciences Institute, USC [2004]).

3. Temple (Temple, Information Sciences Institute, USC [2004]).

4.3.3 Markup Languages

Markup languages such as HTML, XML, XML Schema, and XHTML have become ubiquitous today. Standards for such languages are set by the W3C organization (http://www.w3c.org). There are numerous books describing XML (including its extensions such as XML Schema and XHTML) and applications to electronic commerce (Martin et al., 2000; Schmelzer et al., 2002).

4.3.4 Semantic Web

The network-centric environment for Army knowledge fusion will be based upon the semantic web. Some general introductions are Berners-Lee et al. (2002) and Fensel (2003a, 2003b). RDF (Lassila and Swick, 1999) is a basic ontology language based upon XML. Many of the semantic web languages are based upon RDF. See Patel-Schneider and Fensel (2002) for a discussion of various issues related to the layering of languages that make up the semantic web.

Heflin et al. (2003) discuss the web markup language SHOE (not based on RDF). McGuinness et al. (2003) describe the language DAML-ONT, which is the ontology portion of DAML (DARPA Agent Markup Language). The language OIL is described by Klein et al. (2003). Both OIL and the combination of DAML and OIL (DAML+OIL) are discussed by Fensel et al. (2003b). A more recent language being considered is OWL (Smith, 2003). All of these languages are based on RDF and have been considered by the W3C. OWL is currently a W3C candidate recommendation.

Much work in the semantic web is devoted to the development of ontologies, ontology languages, and support tools for the specification of the information available on web pages and also for appropriately making web services available to automated agents. For discussions of various tools see (Omelayenko et al., 1993; Sure and Studer, 2003; Klein et al., 2003; Engels and Lech, 2003; Sure et al., 2003). Lassila and Adler (1993) discuss how web services (semantically specified) may be utilized by a hand-held (PDA-like) device. A similar device may be utilized by soldiers in the field to access services available on the network (GIG).

4.3.5 Intelligent Search and Retrieval

Maedche et al. (2003) describe techniques for semantically querying and navigating repositories of materials annotated with a semantic markup language. There has also been work on the storage and retrieval of RDF annotated data. See, for example, Davies et al. (2003) and also Broekstra et al. (2003).

Service retrieval (e.g., online web services) is an important topic that is likely to play a major role in the information system needed to support the Army's Future Force. One approach, due to Klein and Bernstein, is based upon the use of an ontology of processes (Klein and Bernstein 2001; Bernstein and Klein, 2002). The work makes use of an approach to modeling processes that has associated with it a number of tools and other applications (Klein and Dellarocas, 2000; Bernstein et al., 1999; Klein, 2003).

4.3.6 Web-Site Management

A tremendous amount of information will be available throughout the GIG for inclusion into the CROP. Automating the process of gathering the relevant data for inclusion is a major issue. There has been some relevant work on web-site management. In particular two systems, Strudel and Tiramisu have been developed (Fernández et al., 1997a, 1997b, 2000; Anderson et al., 1999). These allow the site designer to specify in a declarative fashion what data should be included in the website and how that data should be visually presented.

19

4.4 Advanced Representation and Reasoning

This section covers a number of topics related to representation and reasoning that are likely to be essential to the success of NCW. One is the development of a flexible notion of plans. This has been mentioned earlier. Additionally, it is necessary to recognize the plan of enemy forces given the evidence available. This problem, essential to the vision of NCW, has been investigated under the name of plan recognition, and also it is a specific application of abductive reasoning. Other types of relevant reasoning are logical reasoning such as that found in the literature on logic programming, and also probabilistic or uncertain reasoning. Finally, real-time behavior is likely to be important in many cases, while these reasoning methods often do not exhibit real-time behavior. Work in the area of real-time problem solving addresses the issue of obtaining responses in a time-constrained setting.

4.4.1 Plan Representation

The literature on planning within AI is vast. Additionally, see Allen et al. (1991) for several innovative articles in this area. Additionally, Devanbu and Litman (1996) provide a description logic (CLASP) to represent plans. The notion of subsumption now applies between plan concepts and also between plan concepts and plan instances. This work may be useful in addressing the need for a flexible representation for plans and activities.

GoLog (Levesque et al., 1997; Reiter, 2001) is a high-level programming language for agents based on the situation calculus. The language includes nondeterministic constructs and incorporates deduction into the planning mechanism. Again, the work may prove useful as a basis for a flexible plan representation language.

4.4.2 Plan Recognition

There is a body of literature on plan recognition and more generally on abductive reasoning. Here, rather than constructing a plan and executing it, the goal is to determine on the basis of the available evidence, what is the plan that is being carried out by perhaps the enemy forces. This work will enable the COP to present information on the likely intent and future positions of the enemy forces.

An important early work on plan recognition is put forth by Kautz (1991). A formal theory of plan recognition is developed in which the problem is precisely stated. In this work, the goal of the plan recognition is to produce the most specific plan that accounts for only the observed observations. Therefore, it does not seem amenable to the incorporation of other types of evidence in support of the likelihood of alternative plans.

Charniak and Goldman (1993) present a method of plan recognition in which candidate explanations (i.e., plans) are first retrieved and then are assembled into a Bayesian network. Then the observed actions are entered and Bayesian updating takes place. The result is a probabilistic ranking of candidate plans on the basis of their likelihood in being the actual plan.

Poole (1993) presents a framework for the abduction of Horn clauses, based on the combination of a Bayesian network with Prolog. In Goldman et al. (1999), Poole's work is used as the basis for an approach to plan recognition. These approaches seem to have the flexibility to be extended to allow a variety of types of evidence to affect the likelihood ranking of plans.

Closely related to plan recognition is the abstraction of higher level concepts from temporal data. Shahar (1997) gives a general knowledge-based framework for such a temporal abstraction process. Another related topic is work on finding patterns in temporal data or the dividing of sequences into meaningful episodes (Oates et al., 1997; Cohen, 2001; Cohen and Adams, 2001).

4.4.3 Abduction and Belief Revision

Plan recognition as discussed in the previous section is closely related to the more general topic of abductive reasoning. The problem is to infer what must have been true to account for current observations. Especially relevant is abduction in the context of the occurrence of actions. Some relevant literature is Baral (2000), Poole (1989), and Selman and Levesque (1996).

Josephson and Chandrasekaran (2003) have proposed an architecture for an information system in which abductive inference plays a major role.

Another related area is belief revision (Gärdenfors, 1988, 1992; Goldszmidt and Pearl, 1996; Boutilier, 1998; Friedman and Halpern, 1997). The issue here is how to revise our beliefs given new information. It is often necessary to eliminate beliefs to accommodate the new information and there may be different possible choices that can be made with regard to which beliefs to eliminate.

4.4.4 Logic Programming

The literature on logic programming is quite large. Applications include abduction, nonmonotonic reasoning, planning, reasoning about actions, and various forms of problem solving. Baral and Gelfond (1994) survey the use of logic programming for knowledge representation. The recent book by Baral (2003) covers a newer logic programming language *AnsProlog* which is especially suitable for nonmonotonic reasoning. Reasoning about actions within logic programming languages has been a topic of major interest. Baral, Gelfond, and collaborators have proposed a variety of action description languages with differing capabilities (Baral et al., 1997). An alternative approach is the use of GoLog (Levesque et al., 1997; Reiter, 2001), mentioned earlier.

4.4.5 Uncertain Reasoning

The literature on uncertain or probabilistic reasoning is quite large. Many of the techniques are related to those used in data fusion as discussed in section 4.1. Shafer and Pearl (1990) is a collection of classic papers on the topic of uncertain reasoning in general. Pearl (1998) is a book-length introduction to the combination of graphical models and Bayesian reasoning known as Bayesian Networks. The work of Cowell et al. (1999) and Jensen (1996) are excellent

introductions to probabilistic expert systems. Pearl (2000) is a book devoted to exploring the notion of causality, but in the process lays out the latest thinking of the author on Bayesian networks and probabilistic reasoning in general.

The use of machine learning techniques to learn a type of Bayesian network (*Tree Augmented Naive Bayes*) useful for classification problems is discussed by Friedman et al. (1997). Issues of using Bayesian networks to cluster dynamic processes are considered by Ramoni et al. (2001). These techniques are potentially useful in computing the most probable generating processes for some observed sensor data when that sensor data is continuous. Again, these techniques would supplement those traditionally used in data fusion.

In a number of papers, A. Pfeffer, D. Koller, and collaborators combine Bayesian Networks with the types of more expressive structures found in knowledge representation languages (Koller et al., 1997; Pfeffer et al., 1999; Koller and Pfeffer, 1997, 1998; Pfeffer and Koller, 2000). The result is probabilistic versions of frame-based languages and description logics. This approach is designed to be suitable for representing the battlespace domain. A related approach, but with an added temporal component, has been proposed recently by Sanghai et al. (2003). These methods may turn out to be important in integrating the results of data fusion with higher level reasoning both of a probabilistic and a logical nature.

4.4.6 Real-Time Problem Solving

Because certain computations will have to provide an answer within a limited time period, computing in time-constrained environments is of importance. Unfortunately, many of the reasoning methods discussed earlier do not necessarily have good time-constrained behavior and therefore techniques for performing reasoning with time constraints are likely to be important in the development of the reasoning methods needed to support those aspects of Army knowledge fusion that will demand real-time behavior.

A classic paper is by Boddy and Dean (1994) in which anytime algorithms are discussed. An anytime algorithm will provide an answer at any point in the computation and with more time the quality of the answer will only improve. The composition of real-time modules is discussed by Zilberstein and Russell (1966). Other aspects are considered in Hansen and Zilberstein (2001) and Horvitz (2001). Real-time diagnosis and the construction of plans to discriminate among possible hypotheses are discussed in Ash and Hayes-Roth (1996).

4.5 Contemporary Database Topics

A number of currently active research areas within the field of databases, broadly construed, are potentially quite relevant to the concerns at hand. These include the topics of querying streams and querying unstructured information sources. Also included here are the topics of data mining, web mining, and information retrieval.

4.5.1 Querying Streams

Recently, there has been interest within the field of databases in developing methods for querying streams of continuous data. This work is potentially useful for monitoring sensor data or information on credit card transactions. Much work has been carried out by J. Widom along with students and collaborators at Stanford University (Arasu et al., 2002, 2003; Babcock et al., 2002; Babu and Widom, 2001).

Querying streams of XML data is considered by a number of researchers (Ives et al., 2000b; Gupta and Suciu, 2003). Related issues in routing XML data to the appropriate place are considered by Gupta et al. (2003). See also the Tukwila system mentioned earlier.

The ability to query streams is a critical aspect of Army knowledge fusion. Particular sensors may need to be continuously monitored and alarms sounded (i.e., messages sent to the appropriate people or agents) when significant changes occur. For homeland security, the querying (monitoring) of streams of credit card transactions, information from shipping manifests or passenger lists, also is a useful tool.

4.5.2 Querying Unstructured Information Sources

Developing the capability to process data that are semistructured is an important aspect of current database research. These are data that do not conform to a particular schema as in a database, but are to varying degrees self-describing. Much of the data are in XML format, but it could also be in HTML, or in one of the semantic web languages discussed earlier in section 4.3. It is to be expected that much of the data exchanged over the network supporting NCW will be semistructured, using XML or something similar. Additionally, data available on the web are in this format. The capability of querying such data is crucial.

An excellent introduction to work in this area is the book by Abiteboul et al. (2000). A query language, LOREL, was developed as part of the TSIMMIS Project and is especially designed to query semistructured data. The queries can operate over data having different types and also when some of the data are absent (Abiteboul et al., 1997). This query language was migrated to work on XML (Goldman et al., 2000). Work on developing efficient algorithms for the evaluation of queries over semistructured data has been carried out by Suciu (2002). A general discussion of issues involved in systems developed for querying XML data may be found in Deutsch et al. (1999). Query languages for querying web sites are surveyed by Florescu et al. (1998). These are designed for HTML pages. Querying RDF and RDF Schema is discussed by Broekstra et al. (2003).

4.5.3 Web Mining

Recently there has been a growing interest in web mining; extracting useful information from web pages. Chakrabarti (2003) has written a relatively technical survey of the work on the topic. A less technical work oriented towards business applications is Linoff and Berry (2001).

Knoblock et al. (2002b) apply machine learning techniques to the extraction of data from web pages. Most of this work is dealing with HTML/XML pages, rather than pages annotated in the fashion of the semantic web. If the pages are so annotated, then the mining issues are rather different, and one is able to obtain much more reliable information.

4.5.4 Data Mining

Data mining technology clearly has high relevance to Army knowledge fusion. The literature on data mining is large and growing rapidly. Good surveys of the available techniques in this area include books by Witten and Frank (1999) and Han and Kamber (2001), and an excellent article by Brown (2002) that discusses data mining in the context of work in the area of data fusion.

4.5.5 Information Retrieval

The literature on information retrieval is also vast and cannot be covered here. These techniques are of high importance to Army knowledge fusion because much information is available in textual sources.

One relatively recent and excellent survey is the book by Baeza-Yates and Ribeiro-Neta (1999). Woods (1997) has been working on an interesting approach called Conceptual Indexing. This approach tackles directly the paraphrase problem of different words being used in the texts and in the query. He represents the semantic relationship among concepts, morphological structure, and different words. The resulting taxonomy is then used in retrieval.

4.6 Architectures

The Army is moving toward network-centric systems, pervasive use of sensors, and the introduction of robots into multiple force levels. Hardware and software architectures are essential for systematically structuring systems to support the fusion of these concepts and to achieve the vision of a tightly integrated Future Force made up of humans and machines working together. Large-scale distributed applications using decentralized infrastructures, distributed operating systems, multi-agent architectures, and dynamic adaptability will be required. In addition, architecture technologies for information integration that enable fusion at the application level are needed. Applications will need to interact with a full range of components, ranging from databases, application servers, and data warehouses, to search engines, agents, data streams, and sensors. Classical architecture research is a maturing field, but software architectures designed to address the full range of integration issues required to support the knowledge fusion required for the Future Force are only now emerging.

Shaw (2001) has written an excellent analysis of the evolution and prospects of software architecture research. The JDL data fusion model (White, 1988; Steinberg et al., 1998) describes an architecture that involves multiple agent-like nodes communicating with each other. Software architecture approaches that address the engineering of large, complex software systems include new technologies for component composition and software connector technology for component

interaction (Mehta et al., 2000). In the following sections, we discuss several related topics that pertain to the distributed architecture of GIG.

4.6.1 Agents

The literature on agents is quite large. Only a few relevant pointers can be mentioned here. Subrahmanian et al. (2000) give an excellent survey of much of the work in this area. The remainder of the book covers the IMPACT (Interactive Maryland Platform for Agents Collaborating Together) architecture for building an agent-based system. Bradshaw (1997) is a collection of articles covering various applications of agents.

GoLog (Levesque et al., 1997; Reiter, 2001) is a high-level programming language for agents based on the situation calculus. The idea is that rather than perform planning, an agent can search for a legal execution sequence of a high-level plan/program by reasoning about the effects and preconditions of actions. Knowledge and knowledge-producing actions have been incorporated into the language (Scherl, 2003). ConGolog (De Giacomo et al., 2000) is a concurrent version of GoLog. It includes facilities for prioritizing the execution of concurrent processes, interrupts and exogenous actions.

Lesser and co-authors (Lesser et al., 2000) describe an agent named BIG, designed to gather information over the World Wide Web in support of a decision process. One of its characteristics is the capability to reason about the resource trade-offs of different information gathering approaches.

4.6.2 Grids

Recently there has been work in an area called grid computing. The issue is how to perform computation in a distributed environment where different computational resources and data sources are located on a network (see, for example, Blythe et al., 2003a, 2003b; Deelman et al., 1993). AI planning techniques are used to rapidly generate plans or workflows for solving particular computational problems; given a user specification of the desired result. It is likely that similar techniques may be needed to answer queries over the network established to support the Future Force.

4.6.3 Web Services

Languages developed as part of the semantic web endeavor not only categorize information, but also services (Peer, 2002; Paolucci et al., 2002). Daml-S, the portion of DAML+OIL for describing web services, is discussed by Ankolekar et al. (2002). McIlraith and Son (2002) use Golog for composing web services. Tate et al. (2003) specify agents that plan to achieve goals by utilizing available web services.

4.6.4 Distributed Architectures

Architectures for information integration are emerging as a result of the explosion of the Internet to address the problem of integrating large-scale application systems that must interact with databases, distributed servers, workflow systems, search engines, message queues, Web crawlers, mining and analysis packages, and a large variety of programming interfaces such as ODBC, JDBC, Java objects, SQL, XML, WSDL, SOAP, etc. (Roth et al., 2002). The IRIS project at MIT is developing a novel decentralized infrastructure, based on distributed hash tables that will enable a new generation of large-scale distributed applications (Dabek et al., 2003).

In a distributed setting, processes/agents need to communicate with other processes/agents even if they do not know the location, identity, or even the presence of the others with whom they need to communicate. Techniques have been developed to decouple the producers of information and services from those who consume or use the information or services. One such architecture for accomplishing this decoupling is the publish/subscribe paradigm. It provides for mediators that enable the communication between publishers/producers and clients/subscribers. If a publisher/producer has information or a service, it publishes it via the mediator and the mediator communicates the fact that the information or service is available to those clients who have subscribed to receive notices of the appropriate kind of events. See Eugster et al. (2003) for a survey of work on the publish/subscribe paradigm. An alternate (but related) model for a loosely coupled distributed system based on the concepts of event generation, observation, and notification is discussed by Rosenblum and Wolf (1997).

Minsky and collaborators (Minsky and Ungureanu, 2000; Murata, 2003) have developed a mechanism for regulating such open distributed systems. This method is called Law Governed Interaction (LGI) and provides a scalable method for ensuring that such a system will continue to maintain desired properties even as new agents are added and existing agents are deleted or fail.

5. Analysis and Summary

The construction of information systems that reliably support the Future Force concept depends upon fundamental advances in the field of knowledge fusion. The Future Force concept is not achievable solely using today's technologies. These advances require that the state of the art in the areas identified in this report be advanced and approached from a multidisciplinary perspective. Specifically, it will be necessary to bring these different areas together, to combine the different approaches into a single information system in which different pieces of information are analyzed with a variety of techniques, and the results are fused together into knowledge of the overall picture. Systems that achieve portions of this goal today are developed in an ad hoc manner, making such systems unreliable and unflexible, often unable to scale-up or

adapt to dynamic changes in a real-life military environment. The potential for rapidly introducing specific areas of new fundamental research and combining this research with components ready or near-ready for implementation today is strong.

Currently, representation and reasoning have not played a large role in data fusion research. Going forward, it is essential that these approaches become a major part of knowledge fusion research in order to gain fundamental understanding of the problem and find reliable solutions. This report has considered a variety of languages for representing knowledge and information. In many cases, these languages have been provided with a formally defined syntax and semantics. Reasoning methods and tools have been developed for these languages.

In particular, the following areas are well developed:

1. Information Integration in Databases: How to automatically give a global view to information available in heterogeneous and distributed databases. Techniques have been developed to handle differing schemas and terminology.

2. Ontologies and languages and tools to support their use: How to formally define the meaning of terminologies so that varying terminologies can be manipulated by machines.

3. Semantic Web: The development of languages to make information available on the web (both web pages and web services) interpretable by machines.

This is not to say that work in the above areas is completed. There is still much to be done. But some problems have been isolated, and reasonable solutions have been developed. Work in these areas is mature enough to begin the process of technology transition for some aspects of the knowledge fusion problem.

Integration of the high-level language-oriented reasoning methods and the data oriented mathematical techniques of data fusion remains a fundamental research problem. Some potential useful approaches (pointed out earlier) are the work of D. Brown and collaborators on incorporating world knowledge into techniques for tracking objects and also the work of D. Koller, A. Pfeffer, and others on creating probabilistic versions of a number of knowledge-representation languages. More work is needed on integrating the information obtainable from these different levels.

A great deal of fundamental research in architectures for data fusion remains to be done.

Specific approaches for developing large-scale, distributed architectures that will support the Future Force vision at all force levels needs to be researched. Work in the following areas can be drawn upon:

1. Agents: How to build systems of multiple interacting programs.

2. Grids: How to perform distributed computations.

3. Web Services: Incorporation of services into the semantic web.

4. Distributed Systems: How to create robust distributed architectures.

5. Streaming Data: How to summarize, simplify, identify critical change and critical objects in real-time streams of sensor, video, audio, messaging data.

6. Information Integration: How to design and manage the integration of heterogeneous application components and their interactions.

There are a number of areas of reasoning and representation in which more work is needed for successful knowledge fusion:

1. The determination of the goals of the various human entities (e.g., enemy forces) is a major problem, and the state of current knowledge is not sufficient to easily address the problem. On the basis of available evidence, it is necessary to identify the likely plan of different agents in the battlesphere. More work is needed in plan recognition, situation, impact assessment and awareness, and abductive reasoning. These issues are clearly related to the handling of uncertainty and also the representation of information.

2. It is not clear how to combine the lower level sensor readings and the statistical data fusion predictions with the higher level representation of plans, real-time information from the web and other sources. More work is needed to address this issue both by developing appropriate fusion architectures and new forms of representation, including methods for representing audio and video content.

3. Work is needed on how to properly integrate textual information into the overall CROP. This information may, for example, be reports of human observers.

4. More work is needed on how to handle uncertainty in this setting. Information may be conflicting. It is necessary to manage multiple competing hypotheses and update the probabilities of each based upon newly arriving information.

5. A flexible method of representing actions, decisions, and plans is needed. This includes both the hypothesized actions and plans of the enemy forces and representations to be given to agents or battlespace entities. One of the ideas behind NCW is that commanders do not have to issue top-down detailed instructions, but can rely on the local units to act in accordance with their knowledge to achieve the desired behavior. But the specification (and monitoring) of the action or plan to be given to such a system of knowledgeable actors is an open question.

6. References

Abiteboul, S.; Buneman, P.; Suciu, D. *Data on the Web: From Relations to Semistructured Data and XML*; Morgan Kaufmann Publishers: San Francisco, CA, 2000.

Abiteboul, S.; Quass, D.; McHugh, J.; Widom, J.; Wiener, J. L. The Lorel Query Language for Semistructured Data. *International Journal on Digital Libraries* **1997**, *1* (1), 68–88.

Adali, S.; Emery, R. A Uniform Framework for Integrating Knowledge in Heterogeneous Knowledge Systems. In *Proceedings of the 11th IEEE International Conference on Data Engineering*, Taipai, Taiwan, 6–10 March 1995; IEEE Computer Society Press: New York, 1995; pp 513–520.

Adali, S.; Subrahmanian, V. S. Amalgamating Knowledge Bases, II: Distributed Mediators. *International Journal of Intelligent and Cooperative Information Systems* **1994**, *3* (4), 349–383.

Adali, S.; Subrahmanian, V. S. Amalgamating Knowledge Bases, III: Algorithms, Data Structures, and Query Processing. *Journal of Logic Programming* **1996**, *28* (1), 57–100.

Alberts, D.; Garstka, J.; Stein, F. *Network Centric Warfare: Developing and Leveraging Information Superiority*, CCRP, 2002; pp 2, 68, 118, 126, 127, 133–135, 160, 161.

Allen, J.; Kautz, H.; Pelavin, R.; Tenenberg, J. *Reasoning About Plans*; Morgan Kaufmann Publishers: San Mateo, CA, 1991.

Ambite, J.; Knoblock, C. Flexible and Scalable Cost-Based Query Planning in Mediators: A Transformational Approach. *Artificial Intelligence* **2000**, *118*, 115–161.

Anderson, C.; Levy, A.; Weld, D. Declarative Web-Site Management With Tiramisu. *Proceedings of the International Workshop on the Web and Databases (WebDB)*, Philadelphia, PA, 3–4 June 1999.

Ankolekar, A.; Burstein, M.; Hobbes, J.; Lassila, O.; Martin, D.; McDermott, D.; McIlraith, S.; Narayanan, S.; Paolucci, M.; Payne, T.; Sycara, K. DAML-S: Web Service Description for the Semantic Web. In *The Semantic-Web-ISWC 2002*; Horrocks, I., Hendler, J., Eds.; Springer-Verlag: Berlin, Germany, 2002; pp 348–363.

Antony, R. *Principles of Data Fusion Automation*; Artech House, Inc.: Deedham, MA, 1995.

Arasu, A.; Babcock, B.; Babu, S.; McAlister, J.; Widom, J. Characterizing Memory Requirements for Queries Over Continuous Data Streams. *Proceedings of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, Madison, WI, 2–7 June 2002; pp 221–232.

Arasu, A.; Babu, S.; Widom, J. An Abstract Semantics and Concrete Language for Continuous Queries Over Streams and Relations. *Proceedings of the 9th International Conference on Data Base Programming Languages*, Potsdam, Germany, 6–8 September 2003.

Arenas, M.; Bertossi, L.; Kifer, M. Applications of Annotated Predicate Calculus to Querying Inconsistent Databases. In *Computational Logic – CL2000*, Loyd, J., Dahl, V., Furbach, U., Kerber, M., Lav, K.-K., Palamidessi, C., Moniz Pereira, L., Sagiv, Y., Stuckey, P. J., Eds.; Springer: New York, 2000; LNAI, Vol. 1861, pp 926–941.

Arenas, M.; Bertossi, L.; Chomicki, J. Answer Sets for Consistent Query Answering in Inconsistent Databases. *Theory and Practice of Logic Programming* **2003**, *3* (485), 393–424.

Arens, Y.; Knoblock, C.; Shen, W. Query Reformulation for Dynamic Information Integration. *Journal of Intelligent Information Systems* **1996**, *6*, 99–130.

Arens, Y.; Chee, C.; Hsu, C.; Knoblock, C. Retrieving and Integrating Data From Multiple Information Sources. *International Journal of Intelligent and Cooperative Information Systems* **1993**, *2* (2), 127–158.

Ash, D.; Hayes-Roth, B. Using Action-Based Hierarchies for Real-Time Diagnosis. *Artificial Intelligence* **1996**, *88*, 317–348.

Babcock, B.; Babu, S.; Datar, M.; Motwani, R.; Widom, J. Models and Issues in Data Stream Systems. *Proceedings of the 2002 ACM Symposium on Principles of Database Systems*, Madison, WI, 2–7 June 2002; pp 1–16.

Babu, S.; Widom, J. Continues Queries Over Data Streams. *SIGMOD Record* **2001**, *30* (3), 109–120.

Bader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; Patel-Schneider, P. The Description Logic Handbook: *Theory, Implementation and Applications*; Cambridge University Press: Cambridge, UK, 2003.

Baeza-Yates, R.; Ribiero-Neta, B. *Modern Information Retrieval*; Addison-Wesley-Longman: Essex, England, 1999.

Bar-Shalom, Y.; Fortmann, T. *Tracking and Data Association*; Academic Press: Boston, MA, 1998.

Baral, C. Abductive Reasoning Through Filtering. *Artificial Intelligence* **2000**, *120*, 1–28.

Baral, C. *Knowledge Representation, Reasoning and Declarative Problem Solving*; Cambridge University Press: Cambridge, UK, 2003.

Baral, C.; Gelfond, M. Logic Programming and Knowledge Representation. *Journal of Logic Programming* **1994**, *19*, *20*, 73–148.

Baral, C.; Gelfond, M.; Provetti, A. Representing Actions: Laws, Observations and Hypotheses. *Journal of Logic Programming* **1997**, *31* (1–3), 201–243.

Barker, A.; Brown, D.; Martin, W. Static Data Association With a Terrain-Based Prior Density. *IEEE Transactions on Systems, Man, and Cybernetics* **1998**, *28* (1), 151–157.

Barwise, J. *The Situation in Logic*; CSLI: Stanford, CA, 1989.

Barwise, J.; Seligman, J. *Information Flow*; Cambridge University Press: Cambridge, UK, 1997.

Batini, C.; Lenzerini, M.; Navathe, S. B. A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys* **1986**, *18* (4), 323–364.

Beeri, C.; Levy, A. Y.; Rousset, M. Rewriting Queries Using Views in Description Logics. *Proceedings of the 16th ACM SIGACT-SIGMOD-SIGARTS Symposium on Principles of Database Systems*, Tucson, AZ, 12–14 May 1997.

Berners-Lee, T.; Hendler, J.; Lassila, O. *The Semantic Web: Scientific American*, Special Online Issue, **2002**; 24–30.

Bernstein, A.; Klein, M. Towards High-Precision Service Retrieval. *The Semantic Web – ISWC 2002 Proceedings of the 1st International Semantic Web Conference*, Sardinia, Italy, 9–12 June 2002; Springer-Verlag: Berlin, Germany, 2002, pp 84–101.

Bernstein, A.; Klein, M.; Malone, T. W. The Process Recombinator: A Tool for Generating New Business Process Ideas. *Proceedings of the International Conference on Information Systems*, Charlotte, NC, 12–15 December 1999.

Bertossi, L.; Chomicki, J. *Query Answering in Inconsistent Databases. Logics for Emerging Applications of Databases*; Chomicki, J., Saake, G., van der Meyden, R., Eds.; Springer-Verlag: Berlin, Germany, 2003.

Bertossi, L.; Chomicki, J.; Cortés, A.; Gutiérrez, C. Consistent Answers From Integrated Data Sources in Flexible Query Answering Systems. *Proceedings of the 5th International Conference FQAS*, Copenhagen, Denmark, 27–29 October 2002; Andrewen, T., Motro, A., Christiansen, H., Larsen, H. L., Eds.; Springer-Verlag: Berlin, Germany, 2002, LNAI 2522, pp 71–85.

Blythe, J. Integrating Expectations From Different Sources to Help End Users Acquire Procedural Knowledge. *International Joint Conference on Artificial Intelligence*, Seattle, WA, 9–10 August 2001.

Blythe, J.; Kim, J.; Ramachandran, S.; Gil, Y. An Integrated Environment for Knowledge Acquisition. *International Conference on Intelligent User Interfaces*, Santa Fe, NM, 14–17 January 2001.

Blythe, J.; Deelman, E.; Gil, Y.; Kesselman, C.; Agawal, A.; Mehta, G.; Vahi, K. The Role of Planning in Grid Computing. *Proceedings of the 13th International Conference on Automated Planning and Scheduling, (ICAPS)*, Trento, Italy, 9–13 June 2003a.

Blythe, J.; Deelman, E.; Gil, Y.; Kesselman, C. Transparent Grid Computing: A Knowledge-Based Approach. *Proceedings of the 15th Innovative Applications of Artificial Intelligence Conference (IAAI-03),* Acapulco, Mexico, 12–14 August 2003b.

Boddy, M.; Dean, T. L. Deliberation Scheduling for Problem Solving in Time-Constrained Environments. *Artificial Intelligence* **1994**, *67*, 245–285.

Boutilier, C. A Unified Model of Qualitative Belief Change: A Dynamical Systems Perspective. *Artificial Intelligence* **1998**, *98*, 281–316.

Bovey, B.; Brown, D. Target Tracking Prediction Through Group Correlation Analysis. *Proceedings of the 3rd International Conference on Information Fusion (Fusion 2000)*, Paris, France, 10–13 July 2000; Vol. 1, pp 11–17.

Bradshaw, J. *Software Agents*; AAAI Press: Menlo Park, CA, 1997.

Bravo, L.; Bertossi, L. Logic Programs for Consistently Querying Data Integration Systems. *Proceedings of Practical Aspects of Declarative Languages (PADL03)*, New Orleans, LA, 13–14 January 2003; Dahl, V., Wadler, P., Eds.; Springer: New York, 2003; LNCS 2563, pp 208–222.

Broekstra, J.; Kampman, A.; van Harmelen, F. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 71–89.

Brown, D. E. Data Mining, Data Fusion, and the Future of Systems Engineering. *IEEE International Conference on Systems, Man and Cybernetics*, Washington, DC, 5–8 October 2002; Vol. 1, 26–30.

Brown, D. E. The Regional Crime Analyses Program (ReCAP): A Framework for Mining Data to Catch Criminals. *Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA, 11–14 October 1998; Vol. 3, pp 2848–2853.

Brown, D. E.; Hagen, S. C. Correlation Analysis for Decision Support With Applications to Law Enforcement. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Tokyo, Japan, 12–15 October 1999; Vol. 6, pp 1074–1078.

Brown, D. E.; Oxford, R. Data Mining Time Series With Applications to Crime Analysis. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Tuscon, AZ, 7–19 October 2001; Vols. 3 and 6, pp 1453–1458.

Brown, D. E.; Schaumburg, J. B. A Simulation-Optimization Methodology for Sensor Placement. *Proceedings of the IEEE International Conference on Computational Cybernetics and Simulation*, Orlando, FL, 12–15 October 2001; Vols. 1 and 6, pp 439–443.

Brown, J. S.; Duguid, P. *The Social Life of Information*; Harvard Business School Press: Boston, MA, 2000.

Calvanese, D.; De Giacomo, G.; Lenzerini, M. Structures Objects: Modeling and Reasoning. *The 4th International Conference on Deductive and Object-Oriented Databases (DOOD'95)*, Singapore, 4–7 December 1995; Springer: New York, LNCS 1013, pp 229–246.

Calvanese, D.; De Giacomo, G.; Lenzerini, M. Conjunctive Query Containment in Description Logics With N-Ary Relations. *Proceedings of the 1997 Workshop on Description Logics (DL'97)*, Gif sur Yvette, France, 27–29 September 1997a; pp 5–9.

Calvanese, D.; De Giacomo, G.; Lenzerini, M. Representing and Reasoning on SGML Documents. *Proceedings of the 10th International Symposium on Methodologies for Intelligent Systems (ISMIS'97)*, Charlotte, NC, 15–18 October 1997b.

Calvanese, D.; De Giacomo, G.; Lenzerini, M.; Nardi, D.; Rosati, R. Description Logic Framework for Information Integration. *Proceedings of the 6th International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, Trento, Italy, 20–25 June 1998a; pp 2–13.

Calvanese, D.; De Giacomo, G.; Lenzerini, M.; Nardi, D.; Rosati, R. Source Integration in Data Warehousing. *Proceedings of the 9th International Workshop on Database and Expert Systems Applications (DEXA-98)*, Vienna, Austria, 26–28 August 1998b; pp 192–197.

Calvanese, D.; Lenzerini, M.; Nardi, D. Description Logics for Conceptual Data Modeling. *Logics for Databases and Information Systems*; Chomicki, J., Saake, G., Eds.; Kluwer Academic Publishers: Boston, MA, 1998c; pp 229–264.

Calvanese, D.; De Giacomo, G.; Lenzerini, M. Description Logics: Foundations for Class-Based Knowledge Representation. *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science*, Copenhagen, Denmark, 22–25 July 2002; pp 22–25.

Carvalho, H.; Heinzelman, W.; Murphy, A.; Coelho, C. A General Data Fusion Architecture. *Proceedings of the 6th International Conference on Information Fusion (Fusion 2003)*, Cairns, Queensland, Australia, 8–11 July 2003.

Chakrabarti, S. *Mining the Web: Discovering Knowledge From Hypertext Data*; Morgan Kaufmann Publishers: San Francisco, CA, 2003.

Charniak, E.; Goldman, R. P. A Bayesian Model of Plan Recognition. *Artificial Intelligence* **1993**, (64), 53–89.

Cohen, W. WHIRL: A Word-Based Information Representation Language. *Artificial Intelligence* **2000**, *118*, 288–321.

Cohen, P.; Adams, N. An Algorithm for Segmenting Categorical Time Series Into Meaningful Episodes. *Proceedings of the 4th Symposium on Intelligent Data Analysis*, Springer-Verlag: Berlin, LNCS 2189, 2001; pp 198–207.

Cohen, P.; Shrag, R.; Jones, E.; Pease, A.; Lin, A.; Starr, B.; Gunning, D.; Burke, M. The DARPA High-Performance Knowledge Bases Project. *Artificial Intelligence* **1998**, 25–49.

Cohen, P. Fluent Learning: Elucidating the Structure of Episodes. *Proceedings of the 4th Symposium on Intelligent Data Analysis*, Lisbon, Portugal, 13–14 September 2001; Springer-Verlag: Berlin, LNCS 2189, pp 268–277.

Cowell, R.; Dawid, A.; Lauritzen, S.; Spiegelhalte, D. *Probabilistic Networks and Expert Systems*; Springer: New York, 1999.

Craven, M.; DiPasquo, D.; Freitag, D.; McCallum, A.; Mitchell, T.; Nigam, K.; Slattery, S. Learning to Construct Knowledge Bases From the World Wide Web. *Artificial Intelligence* **2000**, *118*, 69–113.

Cycorp Home Page. http://cyc.com (accessed May 2004).

Dabek, F.; Zhao, B.; Druschel, P.; Stoica, I. Towards a Common API for Structured Peer-to-Peer Overlays. *The 2nd International Workshop on Peer-to-Peer Systems,* Berkeley, CA, 20–21 February 2003.

Davenport, T.; Prusak, L. *Working Knowledge*; Harvard University Press: Cambridge, MA, 1998.

Davies, J.; Weeks, R.; Krohn, V. QuizRDF: Search Technology for the Semantic Web. In *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 133–144.

De Giacomo, G.; Lespérance, Y.; Levesque, H. ConGolog, a Concurrent Programming Language Based on the Situation Calculus. *Artificial Intelligence* **2000**, *121*, 109–169.

Deelman, E.; Blythe, J.; Gil, Y.; Kesselman, C.; Mehta, G.; Vahi, K.; Blackburn, K.; Lazzarini, A.; Arbree, A.; Cavanaugh, R.; Koranda, S. Mapping Abstract Complex Workflows Onto Grid Environments. *Journal of Grid Computing* **2003**, *1* (1).

Department of the Army, Objective Force Task Force. The Objective Force in 2015 White Paper 2002. http://www.afms1.belvoir.army.mil/FASO/white%20Paper(final).pdf (accessed May 2004).

Deutsch, A.; Fernández, M.; Florescu, D.; Levy, A.; Maier, D.; Suciu, D. Querying XML Data. *IEEE Data Engineering Bulletin* **1999**, *22* (3), 10–18.

Devanbu, P.; Litman, D. Taxonomic Plan Reasoning. *Artificial Intelligence* **1996**, *84*, 1–35.

Devlin, K. *Infosense: Turning Information Into Knowledge*; W. H. Freeman and Company: New York, 1999.

Devlin, K. *Logic and Information*; Cambridge University Press: Cambridge, UK, 1991.

Doan, A.; Domingos, P.; Levy, A. Learning Source Descriptions For Data Integration. *Proceedings of the International Workshop on the Web and Databases (WebDB)*, Dallas, TX, 18–19 May 2000.

Doan, A.; HaLevy, A. Efficiently Ordering Query Plans for Data Integration. *Proceedings of the International Conference on Data Engineering (ICDE)*, San Jose, CA, 26 February–1 March 2002.

Doan, A.; Madhavan, J.; Domingos, P.; Halevy, A. Learning to Map Between Ontologies on the Semantic Web. *Proceedings of the International WWW Conference*, Honolulu, HA, 7–11 May 2002.

Draper, D.; HaLevy, A.; Weld, D. The Nimble XML Data Integration System. *Proceedings of the International Conference on Data Engineering (ICDE '01)*, 2001.

Dretske, F. *Knowledge and the Flow of Information*. CSLI Publications: Stanford, CA, 1999.

Duda, R.; Hart, P.; Stork, D. *Pattern Classification*; John Wiley and Sons, Inc.: New York, 2001.

Duschka, O. M.; Genesereth, M. R. Answering Recursive Queries Using Views. In *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, Tucson, AZ, 12–14 May 1997a.

Duschka, O. M.; Genesereth, M. R. Query Planning in Infomaster. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Nagoya, Japan, 23–29 August 1997b; Morgan Kaufmann Publishers: San Francisco, CA.

Duschka, O. M.; Genesereth, M. R.; Levy, A. Recursive Query Plans for Data Integration. *Journal of Logic Programming* **2000**, *43* (1), 49–73.

Engels, R. H. P.; Lech, T. C. Generating Ontologies for the Semantic Web: Onto Builder. *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 91–115.

Etzioni, O. The World Wide Web: Quagmire or Gold Mine? *Communications of the ACM* **1996**, *39* (11), 65–68.

Etzioni, O.; Hanks, S.; Jiang, T.; Karp, R. M.; Madani, O.; Waarts, O. Efficient Information Gathering on the Internet. *Proceedings of Foundations of Computer Science (FOCS96)*, Burlington, VT, 14–16 October 1996.

Etzioni, O.; Weld, D. A Softbot-Based Interface to the Internet. *CACM* **1994**, *37* (7), 72–76.

Eugster, P. T.; Felber, P.; Guerraoui, R.; Kermarrec, A. M. The Many Faces of Publish/Subscribe. *ACM Computing Surveys* **2003**, *35* (2), 114–131.

Falvo, A. Concept Offers Warfighters Battlefield Big Picture; U.S. Joint Forces Command. http://www.jfcom.mil/about/experiments/mco2/crop.htm (accessed 24 May 2004).

Fensel, D. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*; Springer-Verlag: Berlin, Germany, 2001.

Fensel, D.; Hendler, J.; Lieberman, H.; Wahlster, W. *Introduction*. Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003a.

Fensel, D.; van Harmelen, F.; Horrocks, I. OIL and DAML+OIL: Ontology Languages for the Semantic Web. In *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003b; pp 11–31.

Fernández, M.; Florescu, D.; Kang, J.; Levy, A.; Suciu, D. STRUDEL: A Web-Site Management System. *ACM SIGMOD Conference on Management of Data*, Tycsibm, AZ, 13–15 May 1997a.

Fernández, M.; Florescu, D.; Levy, A.; Suciu, D. A Query Language for a Web-Site Management System. *SIGMOD Record* **1997b**, *26* (3).

Fernández, M.; Florescu, D.; Levy, A.; Suciu, D. Declarative Specification of Web Sites With STRUDEL. *Proceedings of the 26th Conference on Very Large Data Bases* (*VLDB 2000*), Cairo, Egypt, 10–14 September 2000.

Florescu, D.; Koller, D.; Levy, A. Using Probabilistic Information in Data Integration. *Proceedings of the 23rd International Conference on Very Large Data Bases (VLDB 1997)*, Athens, Greece, 26–29 August 1997.

Florescu, D.; Levy, A.; Mendelzon, A. Database Techniques for the World Wide Web: A Survey, *SIGMOD Record* **1998**, *27* (3), 59–74.

Friedman, M.; Levy, A.; Millstein, T. Navigational Plans for Data Integration. *Proceedings of the 16th National Conference on Article Intelligence (AAAI-99)*, 18–22 July 1999.

Friedman, M.; Weld, D. S. Efficiently Executing Information-Gathering Plans. *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, Nagoya, Japan, 23–29 August 1997.

Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian Network Classifiers. *Machine Learning* **1997**, *29*, 131–163.

Friedman, N.; Halpern, J. Y. Modeling Belief in Dynamic Systems, Part I: Foundations. *Artificial Intelligence* **1997**, *95*, 257–316.

Garcia-Molina, H.; Papakonstantinou, Y.; Quass, D.; Rajaram, A.; Sagiv, Y.; Ullman, J. D.; Widom, J. The TSIMMIS Approach to Mediation: Data Models and Language*. Journal of Intelligent Information Systems* **1997**, *8* (2), 117–132.

Gärdenfors, P. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*; MIT Press: Cambridge, MA, 1988.

Gärdenfors, P., Ed.; *Belief Revision*; Cambridge University Press: United Kingdom, 1992.

Gelman, A.; Carlin, J.; Stern, H.; Rubin, D. *Bayesian Data Analysis*; Chapman & Hall/CRC: Boca Raton, FL, 2000.

Genesereth, M.; Fikes, R. *Knowledge Interchange Format, Version 3.0 Reference Manual*; Technical Report Logic-92-1; Computer Science Department, Stanford University: Stanford, CA, 1992.

Gil, Y. Knowledge Mobility: Semantics for the Web as a White Knight for Knowledge Based Systems. In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman, H., Whalster, W., Eds.; The MIT Press: Cambridge, MA, 2003; pp 253–278.

Gil, Y.; Ratnakar, V. Trusting Information Sources One Citizen at a Time. *Proceedings of the 1st International Semantic Web Conference (ISWC)*, Sardinia, Italy, 9–12 June 2002.

Goldman, R.; Geib, C.; Miller, C. A New Model of Plan Recognition. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, July 1999.

Goldman, R.; McHugh, J.; Widom, J. From Semistructured Data to XML: Migrating the Lore Data Model and Query Language. *Markup Languages: Theory and Practice* **2000**, *2* (2).

Goldszmidt, M.; Pearl, J. Qualitative Probabilities for Default Reasoning, Belief Revision and Causal Modeling. *Artificial Intelligence* **1996**, *84*, 57–112.

Goodman, I. R.; Mahler, R.; Nguyen, H. T. *Mathematics of Data Fusion*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; p 17.

Gruber, T. R. A Translation Approach to Portable Ontologies. *Knowledge Acquisition* **1993**, *5* (2), 199–220.

Guarino, N. Formal Ontology and Information Systems. *Proceedings of Formal Ontology in Information Systems (FOIS-98)*, Trento, Italy, June 1998; Guarino, N., Ed., Ios Press: Amsterdam, pp 2–15.

Gupta, A.; Suciu, D. Stream Processing of Xpath Queries With Predicates. *SIGMOD*, San Diego, CA, 9–12 June 2003.

Gupta, A.; Suciu, D.; HaLevy, A. The View Selection Problem for XML Content Based Routing. *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2003)*, San Diego, CA, 9–12 June 2003; ACM Press: New York, pp 68–77.

HaLevy, A. Y.; Madhavan, J. Corpus-Based Knowledge Representation. *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, 9–15 August 2003; pp 1567–1572.

Hall, D. *Mathematical Techniques in Multisensor Data Fusion*; Artech House: Dedham, MA, 1992.

Hall, D. L.; Llinas, J. L. *Handbook of Multisensor Data Fusion*; CRC Press, LLC: New York, 2001.

Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufman Publishers: San Francisco, CA, 2001.

Hansen, E.; Zilberstein, S. Monitoring and Control of Anytime Algorithms: A Dynamic Programming Approach. *Artificial Intelligence* **2001**, *126*, 139–157.

Heflin, J.; Hendler, J.; Luke, S. SHOE: A Blueprint for the Semantic Web, In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003, pp 30–63.

Horvitz, E. Principles and Applications of Continual Computation. *Artificial Intelligence* **2001**, *126*, 159–196.

Hull, R. Managing Semantic Heterogeneity in Databases: A Theoretical Perspective. *Proceedings of Conference on Principles of Database Systems (PODS, 1997)*, Tucson, AZ, 1997; pp 51–60.

Information Sciences Institute, University of Southern California, Loom Project Home Page. http://www.isi.edu/isd/LOOM/LOOM-HOME.html (accessed 24 May 2004).

Information Sciences Institute, University of Southern California. Trellis. http://www.isi.edu/ikcap/trellis (accessed 24 May 2004).

Information Sciences Institute, University of Southern California. Expect, A Reflective Architecture for Knowledge Acquisition. http://www.isi.edu/expect (accessed 24 May 2004).

Information Sciences Institute, University of Southern California. Temple. http://www.isi.edu/ikcap/temple (accessed 24 May 2004).

Ives, Z.; Florescu, D.; Friedman, M.; Levy, A.; Weld, D. An Adaptive Query Execution System for Data Integration. *Proceedings of ACM SIGMOD Conference on Management of Data*, Philadelphia, PA, 1–2 June 1999.

Ives, Z.; Forescu, D.; Levy, A.; Friedman, M.; Weld, D. Adaptive Query Processing for Internet Applications. *Data Engineering Bulletin* **2000a**, *23* (2).

Ives, Z.; Levy, A.; Weld, D. *Efficient Evaluation of Regular Path Expressions on Streaming XML Data*; technical report CSE000502; University of Washington: Seattle, WA, 2000b.

Ives, Z.; HaLevy, A.; Weld, D. Integrating Network-Bound XML Data. *Data Engineering Bulletin* **2001**, *24* (2).

Ives, Z.; HaLevy, A.; Weld, D. An XML Query Engine for Network-Bound Data. *The 29th International Conference on Very Large Data Bases (VLDB 2003)*, Berlin, Germany, 9–12 September 2003.

Jensen, F. *An Introduction to Bayesian Networks*; Springer: New York, 1996.

Jhingran, A. D.; Mattos, N.; Pirahesh, H. Information Integration: A Research Agenda. *IBM Systems Journal* **2002**, *41* (4).

Josephson, J. R.; Chandrasekaran, B. Toward a Generic Architecture for Multisource Information Fusion. *Proceedings of Collaborative Technology Alliances Conference. Advanced Decision Architectures*. College Park, MD, 29 April–1 May 2003; pp 111–116.

Karp, P. *The Design Space of Frame Knowledge Representation Systems*; technical note no. 520; SRI AI Center: Menlo Park, CA, 1993.

Karp, P.; Chaudhri, V.; Paley, S. A Collaborative Environment for Authoring Large Knowledge Bases. *Journal of Intelligent Information Systems* **1998**.

Kautz, H. A Formal Theory of Plan Recognition and its Implementation. In *Reasoning About Plans*; Allen, J., Kautz, H., Pelavin, R., Tenenberg, J., Eds.; Morgan Kaufmann Publishers: San Mateo, CA, 1991; pp 70–125.

Kifer, M.; Subrahmanian, V. S. Theory of Generalized Annotated Logic Programming and Its Applications. *Journal of Logic Programming* **1992**, *12* (4), 335–368.

Kingston, J. *High Performance Knowledge Bases: Four Approaches to Knowledge Acquisition, Representation and Reasoning for Workaround Planning*; informatics research report EDI-INF-RR-0050; Division of Informatics: University of Edinburgh, November 2001.

Kirk, T.; Levy, A. Y.; Sagiv, Y.; Srivastava, D. The Information Manifold. In *Working Notes of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, 1995.

Klein, L. *Sensor and Data Fusion: Concepts and Applications*; SPIE Optical Engineering Press: Bellingham, WA, 1993.

Klein, M. A Knowledge-Based Methodology for Designing Reliable Multi-Agent Systems. *Proceedings of the 2nd International Workshop on Software Engineering for Large-Scale Multi-Agent Systems*, Portland, OR, 2003.

Klein, M.; Bernstein, A. Searching for Services on the Semantic Web Using Process Ontologies. *1st Semantic Web Working Symposium (SWWS-1)*, Stanford University, CA, 29 July–1 August 2001.

Klein, M.; Broekstra, J.; Fensel, D.; van Harmelen, F.; Horrocks, I. Ontologies and Schema Languages on the Web. In *Spinning the Semantic Web*, Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003.

Klein, M.; Dellarocas, C. A Knowledge-Based Approach to Handling Exceptions in Workflow Systems. *Journal of Computer Supported Collaborative Work 9* **2000**, (*3/4*).

Klein, M.; Ding, Y.; Fensel, D.; Omelayenko, B. Ontology Management: Storing, Aligning, and Maintaining Ontologies. In *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 47–69.

Knoblock, C.; Lerman, K.; Minton, S.; Musea, I. Accurately and Reliably Extracting Data From the Web: A Machine Learning Approach. *International Journal on Intelligent and Cooperative Information Systems* **2002a**.

Knoblock, C.; Minton, S.; Ambite, J. L.; Ashish, N.; Muslea, I.; Philpot A.; Tejada, S. The Adriadne Approach to Web-Based Information Integration. *International Journal on Intelligent and Cooperative Information Systems* **2002b**.

Knowledge Systems Laboratory, Stanford University. Ontolingua. http://ksl.stanford.edu/software/ontolingua (accessed May 2004).

Koller, D.; Pfeffer, A. Object-Oriented Bayesian Networks. *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UAI, 1997)* Providence, RI, 1–3 August 1997.

Koller, D.; Pfeffer, A. Probabilistic Frame-Based Systems. *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI, 1998)*, Madison, WI, 26–30 July 1998.

Koller, D.; Levy, A.; Pfeffer, A. P-Classic: A Tractable Probabilistic Description Logic. *Proceedings of the AAAI 14th National Conference on Artificial Intelligence*, Providence, RI, 27–31 July 1997.

Kushmerick, N. Wrapper Induction: Efficiency and Expressiveness. *Artificial Intelligence* **2000**, *118*, 15–68.

Lassila, O.; Adler, M. Semantic Gadgets: Ubiquitous Computing Meets the Semantic Web. In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 1993.

Lassila, O.; Swick, R. *Resource Description Framework (RDF) Model and Syntax Specification*; Technical Report; World Wide Web W3C Consortium, 1999— http://www.w3.org/tr/1999/REC-rdf-syntax-19990222 (accessed May 2004).

Lenat, D.; Guha, R. V. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*; Addison-Wesley: Reading, MA, 1989.

Lesser, V.; Horling, B.; Klassner, F.; Raja, A.; Wagner, T.; Zhang, S. BIG: An Agent for Resource-Bounded Information Gathering and Decision Making. *Artificial Intelligence* **2000**, *118*, 197–244.

Levesque, H.; Reiter, R.; Lespérance, Y.; Lin, F.; Scherl, R. GOLOG: A Logic Programming Language for Dynamic Domains. *Journal of Logic Programming* **1997**, *31*, 59–84.

Levy, A. The Information Manifold Approach to Data Integration. *IEEE Intelligent Systems* **1998**, 1313–1316.

Levy, A. Logic-Based Techniques in Data Integration. In *Logic-Based Artificial Intelligence*, Minker, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2000.

Levy, A. Combining Artificial Intelligence and Databases for Data Integration. In *Artificial Intelligence Today; Recent Trends and Developments (Lecture Notes in Artificial Intelligence)*; Woolridge, M., Veloso, M., Eds.; Springer-Verlag: Berlin, Germany, 1999.

Levy, A.; Rousset, M. CARIN: A Representation Language Combining Horn Rules and Description Logics. *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI, 1996)*, Budapest, Hungary, 11–16 August 1996.

Levy, A.; Rousset, M. Combining Horn Rules and Description Logics in CARIN. *Artificial Intelligence* **1998**.

Levy, A.; Weld, D. Intelligent Internet Systems. *Artificial Intelligence* **2000**, *118*, 1–14.

Levy, A.; Rajaraman, A.; Ordille, J. Querying Heterogeneous Information Sources Using Source Descriptions. *Proceedings of the 22nd International Conference on Very Large Databases (VLDB-96)*, Bombay, India, 3–6 September 1996; Vijayaraman, T. M., Buchmann, A. P., Mohan, C., Sarda, N. L., Eds.; Morgan Kaufmann Publishers: San Francisco, 1996a; pp 251–262.

Levy, A.; Rajaraman, A.; Ordille, J. Query-Answering Algorithms for Information Agents. *Proceedings of the 13th National Conference on Artificial Agents*, Portland, OR, 4–8 August 1996b; pp 40–47.

Linoff, G.; Berry, M. *Mining the Web: Transforming Customer Data Into Customer Value*; John Wiley and Sons, Inc.: New York, NY, 2001.

Liu, H.; Brown, D. E. Tracking Multiple Objects in Terrain. *Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA, 11–14 October 1998; Vol. 3, pp 2933–2937.

Lu, J. J.; Moerkotte, G.; Schü, J.; Subrahmanian, V. S. Efficient Maintenance of Materialized Mediated Views. *Proceedings of the International Conference on Management of Data (SIGMOD)*, San Jose, CA, 22–25 May 1995; pp 340–351.

Lu, J. J.; Nerode, A.; Subrahmanian, V. S. Hybrid Knowledge Bases. *IEEE Transactions on Knowledge and Data Engineering* **1996**.

Maedche, A.; Staab, S.; Stojanovic, N.; Studer, R.; Sure, Y. Semantic Portal: The Seal Approach. In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003.

Martin, D.; Birbeck, M.; Kay, M.; Loesser, B.; Pinnock, J.; Livingstone, S.; Stark, P.; Williams, K.; Anderson, R.; Mohr, S.; Baliles, D.; Peat, B.; Ozu, N. *Professional XML*; Wrox Press Ltd.: Birmingham, UK, 2000.

McGuinness, D. Ontologies Come of Age. In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003; 171–195.

McGuinness, D.; Fikes, R.; Stein, L.; Hendler, J. DAML-ONT: An Ontology Language for the Semantic Web. In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003; 65–93.

McIlraith, S.; Son, T. Adapting Golog for Composition of Semantic Web Services. *Proceedings of the 8th International Conference on Knowledge Representation and Reasoning (KR2002)*, April 2002, pp 482–493.

Mehta, N.; Medvidovic, N.; Phadke, S. Towards a Taxonomy of Software Connectors. *Proceedings of the 22nd International Conference on Software Engineering*, Limerick, Ireland, 4–11 June 2000.

Milstein, T.; HaLevy, A.; Friedman, M. Query Containment for Data Integration Systems, *Journal of Computer and System Sciences* **2002**, *66* (1), 20–39.

Minsky, N. H.; Ungureanu, V. Law-Governed Interaction: A Coordination and Control Mechanism for Heterogeneous Distributed Systems. *TOSEM, ACM Transactions on Software Engineering and Methodology* **2000**, *9* (3), 275–305.

Murata, T. On Regulating Open Distributed Systems. Ph.D. Thesis, Rutgers University, NJ, 2003.

Nestorov, S.; Ullman, J.; Wiener, J.; Chawathe, S. Representative Objects: Concise Representations of Semistructured, Hierarchical Data. *Proceedings of the International Conference on Data Engineering (ICDE-97)*, Birmingham, UK, 7–11 April 1997; IEEE Computer Society Press, pp 79–90.

Nie, Z.; Kambhampati, S. Joint Optimization of Cost and Coverage of Query Plans in Data Integration. *Proceedings of ACM CIKM*, Atlanta, GA, 5–10 November 2001.

Nie, Z.; Kambhampati, S. *A Frequency-Based Approach for Mining Coverage Statistics in Data Integration*; technical report ASU CSE TR 03-004; Arizona State University: Tempe, AZ, 2003.

Nie, A.; Nambiar, U.; Vaddi, S.; Kambhampati, S. *Mining Coverage Statistics for Web Source Selection in a Mediator*; technical report ASU CSE TR 02-008; Arizona State University: Tempe, AZ, 2002.

Oates, T.; Schmill, M. D.; Jensen, D.; Cohen, P. R. A Family of Algorithms for Finding Temporal Structure in Data. *Preliminary Papers of the 6th International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, 4–7 January 1997; pp 371–378.

Office of the Army Chief Information Officer/G-6 JWID Scenario http://www.army.mil/ciogb/SAISIO/SAISIOC/swid.scenario.html (accessed 24 May 2004).

Omelayenko, B.; Crubézy, M.; Fensel, D.; Benjamins, R.; Wielinga, B.; Motta, E.; Musen, M.; Ding, Y. UPML: The Language and Tool Support for Making the Semantic Web Come Alive, In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 1993.

Paolucci, M.; Kawamura, T.; Payne, T.; Sycara, K. *The Semantic Web*, *ISWC 2002*; Horrocks, I., Hendler, J., Eds.; Springer-Verlag: Berlin, Germany, 2002; LNCS 2342, pp 333–347.

Papakonstantinou, Y.; Gupta, A.; Garcia-Molina, H.; Ullman, J. A Query Translation Scheme for Rapid Implementation of Wrappers. *Proceedings of the International Conference on Deductive and Object Oriented Databases (DOOD, 1995)*, Singapore, 4–7 December 1995.

Papakonstantinou, Y.; Garcia-Molina, H.; Ullman, J. Medmaker: A Mediation System Based on Declarative Specifications. *Proceedings of the 12th International Conference on Data Engineering*, New Orleans, LA, 26 February–1 March 1996.

Patel-Schneider, P. Fensel, D. Layering the Semantic Web: Problems and Directions. *The Semantic Web*, *ISWC 2002*; Horrocks, I., Hendler, J., Eds.; Springer-Verlag: Berlin, Germany, 2002; LNCS 2342, pp 16–29.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*; Morgan Kaufman Publishers: San Francisco, CA, 1998.

Pearl, J. *Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2000.

Peer, J. Bringing Together Semantic Web and Web Services. *The Semantic Web, ISWC 2002*; Horrocks, I., Hendler, J., Eds.; Springer-Verlag: Berlin, Germany, 2002; LNCS 2342, pp 279–291.

Pfeffer, A.; Koller, D. Semantics and Inference for Recursive Probability Models. *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*, Austin, TX, 30 July–3 August 2000.

Pfeffer, A.; Koller, D.; Milch, B.; Takusagawa, K. SPOOK: A System for Probabilistic Object Oriented Knowledge Representation. *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, Stockholm, Sweden, 30 July–1 August 1999.

Poole, D. Explanation and Prediction: An Architecture for Default and Abductive Reasoning, *Computational Intelligence* **1989**, *5* (1), 97–110.

Poole, D. Probabilistic Horn Abduction and Bayesian Networks. *Artificial Intelligence* **1993**, *64*, 81–129.

Power, C. M.; Brown, D. E. Context-Based Methods for Track Association. *Proceedings of the 5th International Conference on Information Fusion*, 2002; Vol. 2, 1134–1140.

Ramoni, M.; Sebastiani, P.; Cohen, P. Bayesian Clustering by Dynamics. *Machine Learning* **2001**, 1–31.

Reed, S. L.; Lenat, D. B. *Mapping Ontologies Into CYC. Ontologies and the Semantic Web*; AAAI Workshop technical report WS-02-11; AAAI Press: Menlo Park, CA, 2002; pp 63–70.

Reiter, R. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*; The MIT Press: Cambridge, MA, 2001.

Robert, C. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*; Springer: New York, 2001.

Rosenblum, D. S.; Wolf, A. L. A Design Framework for Internet-Scale Observation and Notification. *Proceedings of the 6th European Software Engineering Conference*, Zurich, Switzerland; 22–25 September 1997; Springer-Verlag: Berlin, Germany, 1997; LNCS 1301, pp 344–360.

Roth, M.; Wolfson, D.; Kleewein, J.; Nelin, C. Information Integration: A New Generation of Information Technology. *IBM Systems Journal* **2002**, *41* (4).

Sadjadi, F., Ed. *Selected Papers on Sensor and Data Fusion*; SPIE Optical Engineering Press: Bellingham, WA, 1996.

Sanghai, S.; Domingos, P.; Weld, D. Dynamic Probabilistic Relational Models. *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, 9–15 August 2003; pp 992–997.

Scherl, R.; Levesque, H. Knowledge, Action, and the Frame Problem. *Artificial Intelligence* **2003**, *144*, 1–39.

Schmelzer, R.; Vandersypen, T.; Bloomberg, J.; Siddalingaiah, M.; Hunting, S.; Qualls, M.; Houlding, D.; Darby, C.; Kennedy, D. *XML and Web Services: Unleashed*; Sam's Publishing: Indianapolis, IN, 2002.

Selman, B., Levesque, H. J. Support Set Selection for Abductive and Default Reasoning. *Artificial Intelligence* **1996**, *82*, 259–272.

Shafer, G., Pearl, J., Eds. *Readings in Uncertain Reasoning*; Morgan Kaufmann Publishers: San Mateo, CA, 1990.

Shahar, Y. A Framework for Knowledge-Based Temporal Abstraction. *Artificial Intelligence* **1997**, *90*, 79–133.

Shaw, M. The Coming-of-Age of Software Architecture Research. *Proceedings of the 23rd International Conference on Software Engineering*, Toronto, Ontario, Canada, 12–19 May 2001.

Smith, M.; Welty, C.; McGuiness, D. OWL Web Ontology Language Guide, W3C Candidate Recommendation. *World Wide Consortium*, 18 August 2003; http://www.w3.org /TR/2003/CR-owl-guide-20030818 (accessed May 2004).

Sobiesk, E.; Hamilton, J. A.; Narin, J. A.; Brown, D. E.; Gini, M. Tracking Multiple Objects in Terrain. *IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA, 11–14 October 1998; Vol. 3, pp 2842–2847.

Steinberg, A.; Bowman, C.; White, F. Revisions to the JDL Data Fusion Model. *Proceedings of the 3rd NATO/IRIS Conference*, Quebec City, Quebec, Canada, 19–23 October 1998.

Subrahmanian, V. S. Amalgamating Knowledge Bases. *ACM Transactions on Database Systems* **1994**, *19* (2), 291–331.

Subrahmanian, V. S.; Adali, S.; Brink, A.; Emery, R.; Lu, J. J.; Rajput, A.; Rogers, T. J.; Ross, R.; Ward, C. *HERMES: A Heterogeneous Reasoning and Mediator System*; Technical Report; University of Maryland: College Park, MD, 1995.

Subrahmanian, V. S.; Bonatti, P.; Dix, J.; Eiter, T.; Kraus, S.; Ozcan, F.; Ross, R. *Heterogeneous Agent Systems*; The MIT Press: Cambridge, MA, 2000.

Suciu, D. Distributed Query Evaluation on Semistructured Data. *ACM Transactions on Database Systems* **2002**.

Sure, Y.; Studer, R. A Methodology for Ontology-Based Knowledge Management. In *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 33–46.

Sure, Y.; Erdmann, M.; Studer, R. OntoEndit: Collaborative Engineering of Ontologies. In *Towards the Semantic Web: Ontology-Driven Knowledge Management*; Davies, J., Fensel, D., van Harmelen, F., Eds.; John Wiley and Sons, Ltd.: West Sussex, England, 2003; pp 117–132.

Tate, A.; Dalton, J.; Levine, J.; Nixon, A. Task-Achieving Agents on the World Wide Web, In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003.

Thacker, S.; Seth, A.; Patel, S. Complex Relationships for the Semantic Web, In *Spinning the Semantic Web*; Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Eds.; The MIT Press: Cambridge, MA, 2003; pp 279–315.

Ullman, J. Information Integration Using Logical Views. *Proceedings of the International Conference on Database Theory (ICDT, 1997)*, Delphi, Greece, 8–10 January 1997.

U.S. Army Training and Doctrine Command. The Army Future Force: Decisive 21st Century Landpower. U.S. Army Training and Doctrine pamphlet, August 2002.

U.S. Joint Forces Command, Joint Forces Command Glossary. http://www.jfcom.mil/about/glossary.htm (accessed May 2004).

Waltz, E.; Llinas, J. *Multisensor Data Fusion*; Artech House: Dedham, MA, 1990.

White, F. A Model for Data Fusion. *Proceedings of the 1st National Symposium on Sensor Fusion*, 5–8 April 1998; Vol. 2, 1988.

Wiederhold, G. Mediators in the Architecture of Future Information Systems. *IEEE Computer* **1992**, *25* (3), 38–49.

Wiederhold, G. *Intelligent Integration of Information*; Klumer Academic Publishers: Boston, MA, 2002.

Williams, M. Army Future Force Moves Towards Jointness, *Army AL&T Magazine* **November–December 2003**, pp 42–43.

Witten, I.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques With Java Implementations*; Morgan Kaufman Publishers: Los Altos, CA, 1999.

Woods, W. *Conceptual Indexing: A Better Way to Organize Knowledge*; technical report SMLI TR-97-61; Sun Microsystems Laboratories: Mountain View, CA, April, 1997.

Xue, Y.; Brown, D. E. A Decision Model for Spatial Site Selection by Criminals: A Foundation for Law Enforcement Decision Support. *IEEE Transactions on Systems, Man, and Cybernetics* **2003**, *33* (1), 78–85.

Zilberstein S.; Russell, S. Optimal Composition of Real-Time Systems. *Artificial Intelligence* **1966**, *82*, 181–213.

NO. OF COPIES	ORGANIZATION	NO. OF COPIES	ORGANIZATION
1 (PDF Only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218		ABERDEEN PROVING GROUND
		1	DIR USARL AMSRD ARL CI OK TP (BLDG 4600)
1	COMMANDING GENERAL US ARMY MATERIEL CMD AMCRDA TF 5001 EISENHOWER AVE ALEXANDRIA VA 22333-0001		
1	INST FOR ADVNCD TCHNLGY THE UNIV OF TEXAS AT AUSTIN 3925 W BRAKER LN STE 400 AUSTIN TX 78759-5316		
1	US MILITARY ACADEMY MATH SCI CTR EXCELLENCE MADN MATH THAYER HALL WEST POINT NY 10996-1786		
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS IS R 2800 POWDER MILL RD ADELPHI MD 20783-1197		
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197		
3	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CS IS T 2800 POWDER MILL RD ADELPHI MD 20783-1197		

INTENTIONALLY LEFT BLANK.